MAY 2026 ISSUE

# CIT REVIEW JOURNAL

C | R | J

www.crj.cit.edu.al

INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE

ISSN: 2960-2831 (online)
2788-645X (ISSN-L)

CIT
CANADIAN INSTITUTE
OF TECHNOLOGY

# OPEN YOUR DOOR TO THE WORLD

MAY 2026 ISSUE

CIT
REVIEW
JOURNAL

C | R | J

www.crj.cit.edu.al

# WELCOME MESSAGE

**PROF. DR. ISMAIL KOCAYUSUFOĞLU**
Editor-in-Chief, *CIT Review Journal*
Rector, Canadian Institute of Technology

Dear Readers, Colleagues, and Friends,

It is my pleasure to welcome you to the May 2026 Issue of the CIT Review Journal. This edition brings together a rich collection of research that reflects the rapid technological, economic, and societal transformations shaping our region and the world. The studies featured here explore themes at the intersection of artificial intelligence, digital governance, economic modernization, cybersecurity, and immersive learning technologies, offering timely insights into how innovation continues to reshape institutions, industries, and cultural practices.

Across this issue, our contributors examine the evolving role of AI in public administration, financial systems, labour markets, and cultural analysis. They also address emerging challenges in cybersecurity, digital auditing, and autonomous systems, while highlighting the growing importance of virtual reality and advanced automation in education and engineering. Together, these works illustrate the breadth of inquiry and interdisciplinary collaboration that define our academic community.

I extend my sincere appreciation to all authors, reviewers, and members of the editorial team for their dedication and professionalism in preparing this issue. Your commitment ensures that the CIT Review Journal remains a platform for rigorous scholarship and meaningful dialogue.

To our readers, thank you for your continued engagement and support. I hope the ideas presented in this edition inspire new perspectives and contribute to your research, teaching, and professional practice.

Warm regards,

**Prof. Dr. Ismail Kocayusufoğlu**
Editor-in-Chief, *CIT Review Journal*
Rector, Canadian Institute of Technology

GREETING REMARKS

**DR. MAJLINDA ZHEGU**
Professor of innovation management
Université du Québec à Montréal, Canada

Technological innovation opportunities and challenges

Technological innovation is regarded as one of the pillars of economic growth. It has been understood as a driver of productivity, competitiveness, employment, and industrial transformation. But innovation no longer means merely producing something new, but rather orienting progress toward sustainability, social justice, and the preservation of the conditions of life. This responsibility also requires placing human dignity at the heart of technological choices, so that innovation strengthens freedoms and capacities to act.

Artificial intelligence, like any major technological innovation, can be understood as a pharmakon: it can heal, support, and liberate, but it can also weaken and generate side effects. It can improve access to knowledge, support healthcare, facilitate certain tasks, and open new possibilities for creation and cooperation. Yet it can also increase dependencies, amplify inequalities, weaken critical judgment, or render certain responsibilities invisible. Our responsibility is not to allow AI to develop solely according to the logics of performance, markets, or automation. We must build uses, rules, and institutions capable of preserving autonomy, human dignity, the plurality of knowledge, and the power to act.

In this context, countries such as Albania can become sites of situated innovation, attentive to the real needs of communities, public services, education, health, and social cohesion. Technology can help us live better together if it strengthens civic participation, transparency, intergenerational ties, and solidarity with the diaspora. It can also support sustainable agriculture, territorial protection, cultural memory, and equitable access to resources and opportunities. The challenge, therefore, is to build a distinctive, responsible, and humane technological modernity, capable of serving the common good rather than blindly imitating dominant models.

Warm regards,

**Majlinda Zhegu**
Professor of innovation management
Université du Québec à Montréal, Canada

# DIGITAL AUTOMATION AND LABOUR MARKET RESTRUCTURING IN POST-COVID-19 ALBANIA

***Lisian Roseni*** [1] **, *Reis Mulita*** [2]

*1 University of Tirana, Faculty of Social Sciences, Department of Political Science, Albania*
*2 Canadian Institute of Technology. Faculty of Economy, CIRD Economy, Albania*

## Abstract

The post-COVID-19 pandemic has led to an accelerated pace of digitalization which has transformed the Albanian labour market structurally. This paper analyses the influence of digital transformation and process automation upon employment volume, type of jobs, and required skills, within Albanian private enterprises. Primary research was conducted through a survey of 110 businesses in 12 different economic sectors in 2025. Secondary research was drawn from international bodies including the International Labour Organization, INSTAT, World Bank, and Eurostat. A Digi Score Index, Likert-Scale analysis, Independent Samples T-test, and Pearson correlation were used as methods to assess the influence of digitalization on Albanian labor markets. The study found that new firms created after the onset of the pandemic have shown higher levels of Digital Readiness (Digi Score 64.5% vs. 53.8%). Additionally, there is a marked growth in the demand for employees possessing digitally-based skills; this has been accompanied by a 325% relative increase in the use of flexible employment contracts. Furthermore, it has been observed that six major digital professions categories are missing from Albania's Official National List of Professions, thus creating gaps in regulation and statistics that affect the ability of policymakers to develop accurate policies. The results indicate that digital automation is contributing to the polarization of the labour market based upon education and skill levels, in accordance with theoretical notions related to skill-biased technological change. Finally, the paper includes recommendations aimed at both policymakers, higher education institutions, and business stakeholders.

**Keywords:** Digital Transformation, Labour Market, Employment, Digital Divide, Post-Covid-19

## 1. INTRODUCTION

The Fourth Industrial Revolution has dramatically changed labor markets around the globe by increasing automation and dislocating workers who were employed in routine jobs as well as creating entirely new job categories whose professional development can be difficult for existing institutional arrangements to accommodate (Schumpeter, 1942: 81 - 86; Autor et al., 1998: 1170). COVID-19 accelerated the incremental digitalization of the labour market. As a result, organizations all over the world rapidly adopted remote work, cloud-based services, and digitalized their entire supply chain (George et al., 2020: 1754 - 1756). Albania was no exception, while COVID-19 highlighted the lack of preparedness of the digital readiness of Albania's labor market, it also triggered an irreversible shift in structure in terms of the future of employment. Previous research published in the ICITTBT 2023 proceedings (Roseni, 2023) mapped out the general dimensions of Albania's transformative changes in the digital labor market, employing mainly a scenario planning approach and secondary information. That previous research found a developing skills deficit, emerging policy initiatives, and the absence of detailed data on newly evolving digital professions. This paper makes significant progress beyond this research agenda: it presents original data from a structured survey of 110 Albanian companies; uses statistical methods to test formally hypothesized relationships between variables; and finds an important institutional failure, the non-classification of digital professions, which has previously gone without empirical analysis in the context of Albania. What are the questions that underlie this research? The central research question is: How has digital transformation, defined operationally through process automation, ICT adoptions and new ways of working, reshaped the Albanian labor market since the pandemic?

## 2. LITERATURE REVIEW

The research is guided by two theoretical approaches, each one relating to digitalization. Firstly, according to Schumpeter's (1942: 81-86), the process of "creative destruction" implies that technological innovations will systematically displace existing producers and reallocate labor towards the development of new activities with greater potential for growth. Regarding digitalization in Albania, the ICT-intensive sector has grown from 2.32 percent of all active businesses

*Corresponding author: Lisian Roseni, lisian.roseni@unitir.edu.al

in 2014 to 4.13 percent in 2023 (INSTAT, 2024); conversely, the share of businesses operating in primary sectors has declined. Secondly, based on the skill-biased technological change (SBTC) model proposed by Autor et al. (1998: 1170), information technologies create disproportionate advantages for highly skilled workers, widen pay disparities and cause an increased division of occupations into high-skilled/ high-paid occupations and low-skilled / low paid occupation segments. Therefore, the SBTC model can explain the educational stratification discovered through the DigiScore analysis.

**Digital human** capital was identified by Ognjanović et al. (2024) to be one of the most important factors influencing how well an enterprise performs in terms of output per worker across all western balkans. Blažič and Blažič (2020), identified the importance of digital readiness to enable the effective adoption of productive automation within transitional economies. According to the ILO (2023: 8), ICT workers have been recorded as being some of the fastest growing occupations in the region. However, a major obstacle facing these workers is a lack of formal professional classification and therefore limited institutional recognition and legal protection (Eurofound, 2020; De Stefano, 2016). multilingual corpora (Amel Muminovic, 2025).

## 3. METHODOLOGY

In this paper, we use a combination of both quantitative and qualitative methodologies. Quantitative data has been collected from our own survey, which consisted of a questionnaire that was distributed to private businesses in Albania (approx. 120,000) in 2025. The sampling method used was purposeful, where the 110 businesses selected had a range of sizes. The number of businesses sampled was limited due to resources. There is also an assumption made in the paper that large companies (> 250 employees) and medium-sized companies (50 - 249 employees) will have greater access to digital technologies than small companies (< 50 employees). This is based on the findings of Eurostat (2022), which stated that these two groups account for 82% of all digital investments in Albania and provide 78% of formal employment. There were 110 responses to our survey. Respondents included general managers (61.8 %), marketing specialists (20.9%), HR specialists (5.5%), and finance or IT specialists (8.2%). We assume that respondents provided accurate information regarding their organisation's digital capabilities. The Digital Readiness Score (DigiScore) is a composite measure of four factors including:

1. Technology Infrastructure
2. Process Automation
3. Staff Digital Competences
4. Digital Service Provision

These four components of DigiScore consist of 45 individual questions that are scored on a five point likert scale ranging from "does not apply" (score of 0) through to "fully agree" (score of 5). The scores are then combined to produce an overall percentage score out of 100%. We compared the means of the DigiScore scores for companies established prior to COVID-19 (n=61) and those established since the start of the pandemic (n=46) using independent-samples t tests. A Pearson correlation coefficient was used to examine if there was a linear association between the educational attainment levels of the workforce and DigiScore. Finally, we conducted a regulatory gap analysis by comparing the lists of professions contained in the National List of Professions 2009, the revised list of 2017, and Council of Ministers Decision No. 753/2023.

## 4. DATA ANALYSIS

### 4.1 Digital Readiness: The COVID-19 Dividing Lin

The DigiScore Index indicates that there was a significant digital preparedness divide between businesses prior to the pandemic and those that were formed in response to the pandemic. Businesses that were created prior to COVID-19 had a DigiScore of 53.8%, on average, compared to the average DigiScore of 64.5% for companies developed in response to COVID-19; a difference of 10.7 percentage points. As such, the pandemic has been confirmed as being a structural accelerant rather than merely a one-time impact, as stated by George et al., (2020: 1754-1756).

As it relates to organizational predictors of digital capacity, the educational level of an organization's workforce is the strongest predictor of its ability to adopt technology. There is a negative Pearson correlation between the number of workers having primary/secondary education levels and the DigiScore of $r=-.544$ ($p<.001$) whereas the number of employees holding a bachelors/master's degree is positively correlated with DigiScore at $r=+.528$ ($p<.001$). Furthermore, with 56.9% of all employees across the surveyed organizations possessing no higher than primary/secondary level education, the potential for structural barriers to be present in regards to widespread digital adoption is sizeable. These Pearson correlations empirically support the skill-bias technological change model as described by Autor et al., (1998: 1170), within the Albanian context.

Secondary data also supports this image. Between 2019 and 2023, the percentage of enterprises employing ten or more employees that employed at least one ICT specialist increased from 23.4% to 28.4%. Additionally, between 2019 and 2023, the percentage of employees utilizing computers to perform their job duties increased from 26.8% to 29.9% (INSTAT,

2023: 1-2). Lastly, the wage growth experienced by individuals employed in the ICT industry between 2014 and 2023 averaged approximately 14% per annum compared to the average wage growth experienced by the overall economy of approximately 5.7% per annum (INSTAT, 2024). These figures suggest strong demand for skilled workers capable of adopting digital technologies, as well as increasing labor market polarization.

### 4.2 Employment Volume, Forms, and Automation Demand

Although there has been rapid development in digitalization, the vast majority of full-time jobs remain in place (91.9%) as found in ILO's (2023) research regarding the low rate at which gig economy workers occupy positions in most countries of the Balkans. As part-time employees make up approximately 6.2% of the total workforce of the sampled companies, freelance/short-term contractual employee arrangements comprise about 1.9% of the entire workforce.

There is a noticeable directional trend in the aggregate data provided. For instance, among firms established prior to COVID-19, less than 1% (0.8%) of their employees were employed via flexible agreements (freelance/short term); whereas for firms created either during or after the pandemic, this number increased to nearly 10% (3.4%) - a relative growth rate of 325%.

This directional difference approached statistical significance ($p = .054$), indicating a new form of organizational structure may be developing rather than being due to random chance. This finding was also noted and reported more extensively by Eurofound (2020).

The Likert scale measures indicate a significant gap in attitudes between the two groups of firms regarding the use of digital technologies in recruiting and managing staff (see Table 1). Companies developed post-COVID-19 scored significantly greater on each of the three subscales related to how digital technology impacted their recruitment and HR policies. On average the composite subindex "Digitalization Impact" scores $M=3.63$ ($SD=1.48$) for post-Covid companies, compared to $M=2.38$ ($SD=1.62$) for pre-Covid companies, $t(105)=-4.16$, $p<.001$. The greatest differences were indicated by items assessing whether their staff had sufficient digital knowledge

($\Delta=-1.44, p<.001$) and if they placed greater emphasis on hiring candidates based upon technical competence ($\Delta=-1.48, p<.001$) - this indicates that in the case of companies established in the post-pandemic time frame, digital skills went from being desired to a minimum required level of competencies when hiring.

**Table 1:** *Likert Sub-Index Results: Impact of Digitalisation on Employment Policy (n = 107)*

| Statement (Likert 0–5) | Pre-COVID M ± SD | Post-COVID M ± SD | Δ | p-value |
|---|---|---|---|---|
| Digitalization plays a key role in employment policy | 2.51 ± 1.88 | 3.35 ± 1.97 | −0.84 | 0.027 * |
| Staff possess basic technological knowledge | 2.36 ± 1.60 | 3.80 ± 1.44 | −1.44 | < .001 *** |
| Technical skills are prioritised during recruitment | 2.26 ± 1.68 | 3.74 ± 1.50 | −1.48 | < .001 *** |
| Composite sub-index (aggregate) | 2.38 ± 1.62 | 3.63 ± 1.48 | −1.25 | < .001 *** |

*Note. * $p < .05$; *** $p < .001$. Source: Primary data collected by author, 2025.*

### 4.2 Automation and the Classification Gap

This study spans from 2009 through 2025. It examines how there is an increasing disconnect between developing digital professionals and the process for formally recognizing them. Albania's National List of Professions (NLP) was last amended in 2017. At that time, the number of codes increased from 4,225 to 5,443. Since then, it has not been updated. During and subsequent to the COVID-19 pandemic, the Albanian digital economy have changed. By the end of 2025 at least six consolidated digital professions were missing from the NLP:

(1) Social Media Manager / Digital Marketing Specialist;

(2) Digital Content Creator / Influencer;

(3) Data Scientist / Digital Business Analyst;

(4) Cybersecurity Analyst (as a separate recognized profession from general IT Administration);

(5) Mobile Application Developer / DevOps Engineer;

(6) E-commerce Specialist.

These omissions are significant. According to an INSTAT survey, 86.4% of companies use social media for marketing purposes. Additionally, 75.5% provide products or services online. Moreover, between 2018 and 2022, the percentage of firms employing ICT

specialists rose from 23.4% to 28.4%. Decision No. 753/2023 of the Council of Ministers established fiscal recognition for various broad categories of digital activities (such as software programming, data processing and digital advertising). However, these measures do not establish either professional classification codes, competence profiles, or educational requirements.

The consequence is a paradoxical situation described both by Eurobond (2020) and ETF (2022). Namely, employees who possess recognized vocational qualifications for web development or multimedia (AKAFPK, 2023) have no way to identify themselves as a specific category of worker in national employment statistics because no such NLP code exists. In addition to being an issue in terms of economic inequality, the absence of professional classification provides an institutional basis for the marginalization of workers in the gig-economy, as explained by De Stefano (2016). The lack of clear professional classifications prevents INSTAT from breaking down the increasingly large ICT workforce into sub-professions, thereby hindering targeted, evidence-based policies for reskilling programs. Although digitalization is identified as a priority in the National Employment and Skills Strategy 2023 - 2030, it lacks sufficient detail about labor market demands. Therefore, its programs will continue to be generalized. In addition to these statistical implications for policymaking, the 2024 legal dispute over taxation levels for freelancers under Decision No. 753/2023 indicates that this institutional failure also has significant human implications.

## 5. DISCUSSION

Three related conclusions arise from the results. Firstly, COVID-19 represented an institutional shock that led to increased uptake of digitalization and permanent changes in expectations about workers' roles within organizations, thus supporting the mechanism proposed by George et al. (2020: 1754–1756). The sizeable gap in Digi Score scores between cohorts before the pandemic and after represents a structural, rather than cyclical, divergence in the digital culture of organizations.

Secondly, the labour market appears to be dividing into two distinct groups based on education in line with skill-biased technological change theory (Autor et al., 1998: 1170), as evidenced by the strong negative relationship between primary-educated staff and digital capability and the corresponding positive relationship with tertiary educated staff. In addition to these relationships, 56.9% of employees in the firms surveyed had only primary or secondary qualifications. Gashi & Liça (2023: 20) report that qualified employee recruitment has been a long-standing problem for Albanian SMEs, and this issue is becoming even greater as digital capabilities become required minimums for new hires.

Thirdly, institutions have not adapted sufficiently to changing market conditions. The NLP classification gap represents much more than just an administrative technicality; it represents the mechanisms through which many parts of the expanding digital workforce will remain unseen by the data-collecting tools used to inform statistical and policy decisions. These issues are similar to those identified by ITU (2021) concerning how developing countries experience rapid growth of digital labour markets but also grow too quickly to establish effective regulatory oversight and, similarly, to those reported by ILO (2023: 8) regarding professional misclassification creating barriers to career advancement opportunities and worker protections for individuals employed in occupations adjacent to the information technology sector.In comparison to its regional neighbors, Albania ranks lowest among the Western Balkan countries regarding internet penetration rates at 86.1 percent (World Bank, 2024). Albania ranks behind Kosovo (at 96.6 percent) and near Bosnia and Herzegovina (at 86.7 percent) regarding internet penetration rates; however, Albania's penetration rate is substantially below the EU-wide average of 93.1 percent. As demonstrated by Ognjanović et al. (2024) there exists a significant association between the maturity of a country’s digital infrastructure and an organization’s level of digital capacity in the region. Therefore, the lack of sufficient digital infrastructure in Albania contributes further to the educational and institutional barriers previously mentioned.

## 6. CONCLUSIONS & RECOMMENDATIONS

To begin with, the current study provides a large-scale empirical overview of the impact of the introduction of digital technology on the Albanian labor market. Data was collected through surveys of companies. **The three key findings include:**
a) Acceleration of Digitalization after COVID-19, b) A high degree of structural polarization regarding the digital divide between the better educated and less educated (unskilled), c) An institutional "blind spot" as a result of an outdated classification system for professionals. These findings support the central thesis of this paper: that the changes in Albania's labor market since the COVID-19 pandemic have been due to the country's rapid digital transformation. These changes can be measured by several indicators, including the number of jobs, the form of work, and the size of the digital divide.

**Recommendations for specific stakeholders follow:**

1) Policymakers need to immediately order a third version of the National List of Professions. Specifically, they need to add new codes for the six types of digital professions that were previously classified. This will allow policymakers to develop evidence-based policies related to job creation and job protection based upon their ability to measure employment trends.

2) The National Employment and Skills Strategy 2023 - 2030 needs to be revised so it includes specific digital skill goals that align with NLP Codes. This will enable reskilling programs to target those occupational areas where there is a shortage of qualified personnel as defined by INSTAT data.

3) Institutions of higher education and vocational schools need to provide digital literacy training outside of major cities and narrow the significant digital literacy disparities that exist between urban and rural residents, which are documented extensively (e.g., World Bank, 2022; p. 25) and continue to grow.

4) Large national/international corporations need to invest in formal in-house training of low-skilled/ low-educational attainment employees. Low-skilled, low-educational attainment employees comprise approximately two-thirds of the total workforce but are disproportionately vulnerable to technological unemployment.

5) Future studies should use an analytical approach that examines both company and individual worker levels. Future studies should examine the long-term financial stability and contract security of workers employed in unclassified digital occupations, along with demographic information about workers in unclassified digital occupations.

## 8. REFERENCES

Agjencia Kombëtare e Arsimit, Formimit Profesional dhe Kualifikimeve [AKAFPK]. (2023). Katalogu i kualifikimeve profesionale. AKAFPK.

Autor, D. H., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: Have computers changed the labor market? The Quarterly Journal of Economics, 113(4), 1169–1213. https://doi.org/10.1162/003355398555874

Banka Botërore [World Bank]. (2022). Republic of Albania: Improving equitable access to high standard public services through GovTech (p. 25). World Bank Group.

Blažič, B. J., & Blažič, A. J. (2020). Overcoming the digital-divide puzzle in the 21st century. Information Technology and People, 33(2), 885–914.

De Stefano, V. (2016). The rise of the "just-in-time workforce": On-demand work, crowd work, and labor protection in the gig economy. International Labour Review, 155(3), 471–490.

ETF – European Training Foundation. (2022). Digital skills and education in the Western Balkans. ETF.

Eurofound. (2020). New forms of employment: 2020 update. Publications Office of the European Union.

Eurostat. (2022). Digital economy and society statistics, enterprises. Eurostat.

Gashi, A., & Liça, S. (2023). SME performance constraints in Albania. Albanian Journal of Business and Management, 5(1), 15–28.

George, G., Lakhani, K. R., & Puranam, P. (2020). What has changed? The impact of COVID pandemic on the technology and innovation management research agenda. Journal of Management Studies, 57(8), 1754–1756. https://doi.org/10.1111/joms.12634

ILO - International Labour Organization. (2023). Digital employment and skills development in Albania (pp. 8–10). ILO.

INSTAT. (2023). Anketa e TIK-ut në ndërmarrje 2023 (pp. 1–2). INSTAT.

INSTAT. (2024). Të dhëna sekondare mbi pagat dhe ndërmarrjet aktive 2014–2023. INSTAT.

ITU - International Telecommunication Union. (2021). Digital skills assessment for the ICT sector in Albania. ITU.

Ognjanović, J., Đorđević, B., & Janković-Milić, V. (2024). Digital human capital and enterprise performance in the Western Balkans. International Journal of Information Management, 74, 102–115.

Roseni, L. (2023). Digital transformation of the Albanian labour market. In Proceedings of the 3rd International Conference on Intelligence-Based Transformations of Technology and Business Trends (ICITTBT 2023) (pp. 167–171). Canadian Institute of Technology.

Schumpeter, J. A. (1942). Capitalism, socialism and democracy (pp. 81–86). Harper & Brothers.

VKM Nr. 627, datë 11.6.2009 – Lista Kombëtare e Profesioneve. Këshilli i Ministrave të Shqipërisë.

VKM Nr. 514, datë 20.9.2017 – Rishikimi i Listës Kombëtare të Profesioneve. Këshilli i Ministrave të Shqipërisë.

VKM Nr. 753, datë 20.12.2023 – Dispozita zbatuese të tatimit mbi të ardhurat. Këshilli i Ministrave të Shqipërisë.

World Bank. (2024). World Development Indicators: Internet penetration (% of population). World Bank Group.

# AI-BASED THEMATIC AND SEMANTIC ANALYSIS OF KORÇA TRADITIONAL SONGS: A COMPUTATIONAL FOLKLORISTICS APPROACH

***Gloria Tyxhari*** [1]

*1 Department of Statistics and Applied Informatics, Faculty of Economy, University of Tirana, Albania*

## Abstract

The Korça serenade is one of the most distinctive urban folk traditions in Albania, combining lyrical poetry with a characteristic vocal style transmitted orally across generations. Despite its cultural importance, no computational study has examined its textual corpus to date. This paper presents the first AI-assisted analysis of Korça traditional song lyrics, applying a multi-method natural language processing pipeline to a verified dataset of 15 songs. The methodology is implemented in Python and combines keyword-based thematic and sentiment scoring, Term Frequency Inverse Document Frequency vectorization using scikit-learn, hierarchical clustering with Ward linkage, principal component analysis, and semantic co-occurrence network construction. Results reveal that love and separation together account for 67% of the corpus, with emotionally mixed tonality in 60 percent of songs. The co-occurrence network exposes a central semantic triad connecting heart, fire, and love as the lexical core of the tradition.

**Keywords:** AI; Natural Language Processing; Tf-Idf; Clustering; Albanian Language; Cultural Heritage; Korça Serenade

## 1. INTRODUCTION

The preservation of intangible cultural heritage has become a priority in digital humanities following the UNESCO Convention for the Safeguarding of Intangible Cultural Heritage of 2003, which recognized the importance of documenting oral traditions and traditional music (Hou et al., 2022). In Albania, several musical traditions remain under-documented in computational terms despite their recognized cultural significance.

The Korça serenade, locally known as Serenata Korçare, is a lyric-centered urban folk tradition originating in the late nineteenth century, registered in the Albanian National Inventory of Intangible Cultural Heritage (Albanian ICH Register, 2022). The songs address love, separation, nostalgia for homeland, and the passage of time, encoding a cultural sensibility transmitted across generations.

No computational study has previously examined this tradition at the textual level. This work presents the first NLP-based analysis of Korça song lyrics, producing a baseline characterization that can serve as a reference for future studies. The paper extends the author's prior research on machine learning classification (Tyxhari and Vika, 2023) and synthetic data methods (Tyxhari and Martiri, 2022).

## 2. RELATED WORK

### 2.1 Computational Analysis of Folk Song Lyrics

Rosillo-Rodes et al. (2024) applied NLP and machine learning to over 2,000 Flamenco lyrics using a Multinomial Naive Bayes classifier and a co-occurrence network, providing the closest methodological precedent to the present study. Huang et al. (2022) employed recurrent neural networks for data mining of Huaer folk lyrics, a Chinese intangible heritage tradition. Han et al. (2025) proposed automatic motive segmentation in Korean folk songs. Pratama et al. (2024) applied TF-IDF weighting to classify Indonesian folk songs into regional categories, establishing a direct precedent for the TF-IDF approach used here

### 2.2 NLP for the Albanian Language

Albanian is a low-resource language with scarce annotated corpora. Biba and Mane (2014) presented the first sentiment analysis approach for Albanian. Kadriu et al. (2022) contributed a human-annotated dataset of over 10,000 social media comments. Nuci et al. (2024) demonstrated RoBERTa fine-tuning for Albanian text at LREC-COLING 2024. Misini et al. (2024) addressed authorship attribution in Albanian texts achieving F1 scores above 0.90. The present work uses a language-agnostic statistical approach requiring no pre-trained resources.

*Corresponding author: Gloria Tyxhari, gloria.tuxhari@unitir.edu.al

## 3. DATASET AND PREPROCESSING

The dataset consists of 15 Korça traditional songs collected from a publicly available community-maintained archive of Albanian folk song lyrics, verified against known recordings and cross-referenced with printed collections. One entry, "Qeraxhiu i Grebenesë", is explicitly marked as non-serenade in the source archive and is retained to test whether the computational pipeline recovers this editorial distinction.Preprocessing was implemented in Python 3.12. Each text was lowercased and non-alphabetic characters were removed with a regular expression substitution. A custom stop word list of 68 tokens was constructed, covering Albanian function words and vocative filler terms characteristic of folk song address, such as "moj", "oh", and "ah". Tokens shorter than four characters were excluded. No stemming or lemmatization was applied, as no reliable Albanian stemmer exists and the small vocabulary does not introduce significant inflectional noise into TF-IDF weighting.

## 4. EXPERIMENTAL PROTOCOL

Table 1 summarizes the five-step pipeline, the configuration of each method, and its output. The libraries used are scikit-learn 1.4 for TF-IDF vectorization, KMeans initialization, and PCA; scipy 1.13 for hierarchical clustering and dendrogram rendering; NetworkX 3.3 for co-occurrence graph construction; and matplotlib 3.9 for visualization. All random states were fixed at seed 42 for reproducibility.

| Step | Method | Configuration | Output |
|---|---|---|---|
| 1 | Keyword scoring | Thematic lexicon 8 categories; polarity dicts 11+11 terms; intensity = emotional tokens / song length x 40, capped at 5 | Theme, polarity, intensity (1-5) |
| 2 | TF-IDF | TfidfVectorizer: max_features=200, min_df=1 | 15 x 200 feature matrix |
| 3 | Hier. | Ward linkage, Euclidean distance, k=4 | 4 clusters; dendrogram |
| 4 | PCA | n_components=2, random_state=42 | 2D projection; explained variance |
| 5 | Co-occ. network | Window size=5; top 22 nodes; top 35 edges by weight | Weighted undirected graph |

**Table 1:** *Five-step experimental pipeline with full configuration.*

### 4.1 STEP 1: KEYWORD-BASED SCORING

A domain-specific thematic lexicon was manually constructed with eight categories and Albanian lemma lists drawn from the corpus vocabulary. Sentiment polarity was determined by comparing positive and negative term counts per song, with a tolerance margin of one to handle mixed-register songs. Emotional intensity was computed as:

$$intensity = min(5, max(1, round((n\_pos + n\_neg) / n\_words \times 40)))$$

where n_pos and n_neg are the counts of matched positive and negative lemmas respectively and n_words is the token count after preprocessing

### 4.2 STEP 2: TF-IDF VECTORIZATION

TF-IDF vectorization was applied using sklearn.feature_extraction.text.TfidfVectorizer with max_features=200 and min_df=1, producing a 15 x 200 matrix. TF-IDF rewards terms distinctive to individual songs while penalizing terms appearing uniformly across the corpus, making it appropriate for identifying the lexical fingerprint of each song.

### 4.3 STEPS 3 AND 4: CLUSTERING AND PROJECTION

Agglomerative hierarchical clustering was applied via scipy.cluster.hierarchy.linkage with method='ward' and metric='euclidean'. Ward linkage minimizes total within-cluster variance at each merge, producing compact and interpretable clusters

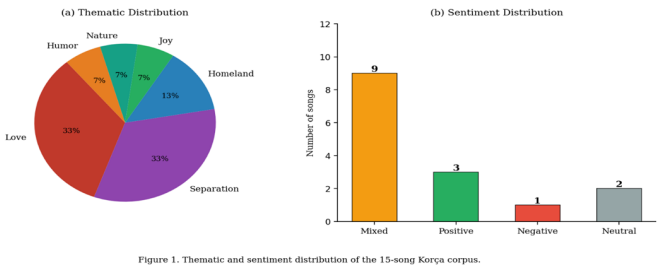



**Figure 1:** *a) Thematic distribution. (b) Sentiment polarity distribution across the 15-song corpus.*

The number of clusters k=4 was chosen by visual inspection of the dendrogram. PCA was applied using sklearn.decomposition.PCA with n_components=2 to project songs from the 200-dimensional TF-IDF space to two dimensions for visualization.

### 4.4 STEP 5: CO-OCCURRENCE NETWORK

A weighted undirected graph was constructed using NetworkX. A sliding window of size five was applied to the preprocessed token sequence of each song. For each pair of tokens within the window, an edge was added or its weight incremented. The graph was restricted to the 22 most frequent content words as nodes and the 35 highest-weight pairs as edges. Node size is proportional to degree centrality and edge width to co-occurrence weight.

## 5. RESULTS AND DISCUSSION

### 5.1 CORPUS STATISTICS AND THEMATIC DISTRIBUTION

Table 2 summarizes the corpus statistics. Figure 1 shows the thematic and sentiment distributions. Love and separation each account for five songs (33.3 percent each), for a combined total of 67 percent, consistent with the known characteristics of the Korça serenade as a tradition centered on romantic expression and longing. The single humor entry, “Shtëpinë moj do ta shes”, is structurally distinct in its playful register and its reference to café culture.

| Metric | Value | Notes |
|---|---|---|
| Total songs | 15 | Verified against recordings |
| Total tokens | 864 | After preprocessing |
| Vocabulary size | 412 | Unique tokens |
| Mean length | 57.6 words | Std. dev. 22.3 words |
| Shortest song | 20 words | Kur vjen behari |
| Longest song | 98 words | Himni i Dashurisë |
| Dominant theme | Love/Separation | 67% of corpus (10/15) |
| Dominant emotion | Mixed | 60% of corpus (9/15) |
| Mean intensity | 1.80 / 5 | Restrained expression |

**Table 2:** *Corpus statistics and overall results.*

The sentiment analysis reveals that 60 percent of songs carry a mixed emotional tone, combining positive and negative registers within the same lyrical unit. This prevalence of emotional ambivalence distinguishes the Korça style from more polarized folk traditions. The mean emotional intensity of 1.80 on a five-point scale confirms that the tradition favors restrained expression over dramatic intensification

### 5.2 TF-IDF LEXICAL FINGERPRINTS

The TF-IDF analysis reveals that each song possesses a distinctive lexical identity recoverable by automated means without manual annotation. The term nënë (mother) is the most discriminative feature of Syçkat si ato të tuat, reflecting its unusual address to a parental figure rather than a romantic partner. The term ndamë (we separated) is the top feature of Himni i Dashurisë. The term harroje (forget) is exclusive to Harroje dashurinë tonë, while zemrat (hearts) and naze (graces) are the fingerprint of O moj Korçare. These results confirm that TF-IDF can recover culturally meaningful distinctions in a low-resource folk song corpus with a vocabulary of 412 unique tokens.

### 5.3 CLUSTER ANALYSIS

The hierarchical clustering identifies four semantically coherent groups summarized in Table 3. Cluster A contains songs with positive or neutral register and lexical fields related to seasons and social scenes. Cluster B clusters songs characterized by guitar and window imagery consistent with the performative context of the serenade. Cluster C clusters songs dominated by the heart-fire-love semantic triad. Cluster D contains songs with the highest negative scores and vocabulary centered on departure and waiting.

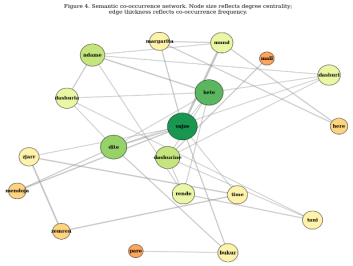


**Figure 2:** *Semantic co-occurrence network. Node size reflects degree centrality; edge thickness reflects co-occurrence frequency.*

Qeraxhiu i Grebenese falls in an isolated position in both the dendrogram and the PCA scatter, confirming that its non-serenade status as marked in the source archive is empirically supported by the computational analysis without any manual intervention the future studies. Together, these dimensions and sub-dimensions provide a structured approach of a core framework, for organizations in the telecommunications industry to assess their Big Data maturity and strategically improve their capabilities to further improve their performance and decision-making.

| Cluster | Dominant theme | Emotion |
|---|---|---|
| A: Nature/Lightness | Nature / Humor | Positive / Neutral |
| B: Lyric Serenade | Love | Mixed / Positive |
| C: Classic Sevdaja | Love / Separation | Mixed |
| D: Melancholy/Sep. | Separation | Negative / Mixed |

**Table 3:** *Cluster membership and dominant characteristics.*

### 5.4 SEMANTIC CO-OCCURRENCE NETWORK

Figure 2 shows the co-occurrence network. Table 4 lists the six most frequent word pairs. The network reveals that vajzë (girl) is the highest-degree node with degree 8, followed by këtë (this/here) with degree 7 and ditë (day) with degree 6. The triad connecting zemrën (heart), zjarr (fire), and dashurinë (love) is the most densely connected subgraph, with a combined co-occurrence weight of five for the zemrën-zjarr pair alone, more than double any competing dyad outside this triad. This triplet constitutes thë semantic core of the Korça lyrical tradition.

A secondary cluster connects mall (longing) and Selanik (Thessaloniki), reflecting the historical reality of Albanian economic migration and its encoding as a geographical separation motif. The peripheral position of the humor cluster, anchored by doktore and kafene, is consistent with the outlier status of Shtepinë moj do ta shes identified across all analyses.

| Word pair | Co-occ. freq. | Interpretation |
|---|---|---|
| vajzë, mund | 8 | Girl / impossibility: unrequited longing motif |
| zemrën, zjarr | 5 | Heart / fire: central semantic triad |
| time, zemrën | 4 | My / heart: first-person lyric address |
| time, zjarr | 4 | My / fire: emotional self-description |
| këtë, ndamë | 4 | This / we-separated: separation motif |
| dashurinë, mall | 3 | Love / longing: defining thematic pair |

**Table 4:** *Top co-occurring word pairs with cultural interpretation.*

## 6. CONCLUSION

This paper presents the first computational analysis of Korça traditional song lyrics, applying a five-step NLP pipeline implemented in Python 3.12 to a verified dataset of 15 songs. Results are consistent across all methods: the tradition is characterized by the thematic dominance of love and separation (67 percent), a preference for emotionally mixed and restrained expression (60 percent mixed, mean intensity 1.80/5), and a semantic core organized around the triad of heart, fire, and love. TF-IDF analysis demonstrates that individual songs maintain distinctive lexical identities despite thematic convergence. Hierarchical clustering and PCA reveal four interpretable semantic groups that align with culturally recognizable categories. The co-occurrence network confirms the outlier status of the single non-serenade entry without requiring manual labels.

The main limitation is dataset size, which restricts statistical robustness and prevents supervised classification. Future work should expand the corpus, incorporate audio feature extraction for multimodal analysis, and conduct cross-regional comparison with other Albanian folk traditions. The dataset and full source code are made available as companions to this paper.

**Acknowledgements**

The author thanks the maintainers of the online Albanian folk song archives whose collections made this dataset possible.

## 7. REFERENCES

Albanian Intangible Cultural Heritage Register. (2022). Serenata Korcare. Ministry of Culture of Albania. Retrieved from https://regjistritkj.al/en/serenata-korcare/

Biba, M., & Mane, M. (2014). Sentiment analysis through machine learning: An experimental evaluation for Albanian. ResearchGate.

Han, D., Jeong, D., & Nam, J. (2025). Motive-level analysis of form-functions association in Korean folk song. arXiv:2508.10472.

Hou, Y., Kenderdine, S., Picca, D., Egloff, M., & Adamou, A. (2022). Digitizing intangible cultural heritage embodied: State of the art. Journal on Computing and Cultural Heritage, 15(3), 55. https://doi.org/10.1145/3494837

Huang, Y., et al. (2022). Intangible cultural heritage management using machine learning: A case study of Northwest folk song Huaer. Scientific Programming. https://doi.org/10.1155/2022/1383520

Kadriu, F., Murtezaj, D., Gashi, F., Ahmedi, L., Kurti, A., & Kastrati, Z. (2022). Human-annotated dataset for social media sentiment analysis for Albanian language. Data in Brief, 43, 108436.

Martiri, E., Kalemi, E., & Tyxhari, G. (2014). A formal object definition in categorical shaping. Academic Journal of Interdisciplinary Studies, 3(1), 427-431.

Misini, A., Canhasi, E., Kadriu, A., Fetahi, E., & Afzaal, M. (2024). Automatic authorship attribution in Albanian texts. PLOS ONE. https://doi.org/10.1371/journal.pone.0310057

Nuci, K. P., Landes, P., & Di Eugenio, B. (2024). RoBERTa low resource fine tuning for sentiment analysis in Albanian. Proceedings of LREC-COLING 2024. Torino, Italia: ELRA and ICCL.

Pratama, et al. (2024). The role of automated classification in preserving Indonesian folk and national songs. Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-031-60012-8_18

Rosillo-Rodes, P., San Miguel, M., & Sanchez, D. (2024). Computational lexical analysis of Flamenco genres. arXiv:2405.05723.

Tyxhari, G., & Martiri, E. (2022). A systematic review of synthetic data generation methods. Circular Economy, 433.

Tyxhari, G., & Vika, B. (2023). Kontributi i algoritmave te machine learning ne klasifikim. Revista Shqiptare Social Ekonomike, 110, 37-47.

# THE ROLE OF E-COMMERCE IN BUSINESS PERFORMANCE: A COMPARATIVE STUDY WITH TRADITIONAL BUSINESS MODELS IN ALBANIA

***Besjana Mema*** [1]

*1 Department of Information Tecnology, Mediterranean University of Albania, Albania*

## Abstract

This study examines the role of e-commerce in modern business and evaluates its importance compared to traditional business models, with a focus on Albania. As digital transformation accelerates globally, businesses increasingly adopt e-commerce to enhance efficiency, expand market reach, and improve customer engagement. A quantitative research design was employed, involving a sample of 200 businesses operating in online, traditional, and hybrid models. The study investigates adoption levels, perceived advantages, challenges, and performance outcomes. Findings indicate that e-commerce adoption significantly improves business performance, particularly in terms of cost efficiency, customer reach, and competitiveness. However, barriers such as technological limitations, security concerns, and infrastructural constraints continue to hinder adoption. The study highlights the importance of supportive policies and digital infrastructure development in emerging economies. The results provide practical implications for businesses and policymakers aiming to foster digital transformation.

**Keywords:** E-Commerce, Digital Transformation, Business Performance, Traditional Business, Albania

## 1. INTRODUCTION

The rapid evolution of the global economy has fundamentally transformed the way businesses operate, compete, and interact with consumers. Among the most significant drivers of this transformation is the rise of electronic commerce, commonly referred to as e-commerce. Over the past two decades, advancements in information and communication technologies have enabled businesses to move beyond the constraints of physical marketplaces and embrace digital platforms that facilitate transactions in real time and across geographic boundaries. This shift has not only redefined commercial practices but has also altered the expectations and behaviors of consumers, who increasingly demand convenience, speed, and personalized experiences in their interactions with businesses. E-commerce has evolved from a supplementary channel of distribution into a central pillar of modern business strategy. While traditional business models have historically relied on physical presence, face-to-face interactions, and localized markets, digital platforms now allow firms to scale operations rapidly and access global audiences with relatively low marginal costs. This transition has introduced a new paradigm in which competitiveness is no longer determined solely by location or physical assets, but by technological capability, data utilization, and the ability to respond quickly to market dynamics. In this sense, e-commerce represents not just a technological innovation, but a structural transformation in how value is created and delivered in the contemporary economy.

From a broader perspective, the expansion of e-commerce has been closely linked to global digitalization processes that have accelerated in recent years, particularly during disruptive events such as the COVID-19 pandemic. The restrictions imposed on physical movement and traditional retail operations forced businesses and consumers alike to adopt digital solutions at an unprecedented pace. As a result, many firms that had previously been hesitant to engage in online commerce were compelled to integrate digital tools into their operations. This period demonstrated that e-commerce is not merely an optional business strategy, but an essential mechanism for resilience and continuity in times of crisis. Moreover, it revealed that digital readiness can significantly influence a firm's ability to adapt and survive in volatile environments. Despite the evident advantages of e-commerce, its adoption is not uniform across regions and industries. In developing economies, including Albania, the transition toward digital business models remains incomplete and often fragmented. While urban centers such as Tirana have experienced notable growth in internet penetration and digital engagement, many businesses continue to rely on traditional practices due to limitations in infrastructure, financial resources, and technological expertise. Additionally, cultural factors such as trust in online transactions and preference for in-person interactions continue to influence consumer behavior, slowing the pace of digital transformation. These challenges highlight the existence of a digital divide that affects both businesses and consumers, creating disparities in access to the opportunities offered by e-commerce.

*Corresponding author: Besjana Mema, *besjana.mema@umsh.edu.al*

In my view, the coexistence of traditional and digital business models should not be interpreted solely as a transitional phase, but rather as a complex and dynamic interaction where each model offers distinct advantages. Traditional businesses retain value in their ability to provide tangible experiences and build personal relationships with customers, which are particularly important in markets where trust plays a critical role. However, the limitations of physical operations—such as higher costs, restricted reach, and lower scalability—make it increasingly difficult for such models to compete in a digitalized environment. On the other hand, e-commerce offers unparalleled opportunities for efficiency, innovation, and market expansion, but it also introduces new risks related to cybersecurity, data privacy, and technological dependency. Therefore, the future of business is likely to be characterized by hybrid models that integrate the strengths of both approaches while mitigating their respective weaknesses. Given this context, understanding the role of e-commerce in shaping business performance and competitiveness is of critical importance. This study aims to explore how e-commerce is being adopted and utilized by businesses, particularly in the Albanian context, and to evaluate its impact compared to traditional business models. By examining both the advantages and the challenges associated with digital transformation, the research seeks to provide a comprehensive perspective on the evolving business landscape. Ultimately, the findings are intended to contribute not only to academic discourse but also to practical decision-making by businesses and policymakers striving to navigate the complexities of the digital economy.

## 2. LITERATURE REVIEW

The concept of e-commerce has evolved significantly over time, with early definitions emphasizing its technological foundation and more recent interpretations highlighting its strategic and economic importance. One of the earliest definitions was provided by Clarke (2000), who described e-commerce as the conduct of commercial activities in goods and services with the assistance of telecommunications-based tools. (rogerclarke.com) This early perspective reflects the initial stage of e-commerce development, where the focus was primarily on electronic transactions and communication technologies. As the internet expanded in the late 1990s and early 2000s, e-commerce became increasingly associated with web-based transactions and online marketplaces, marking a shift toward digital platforms as the core environment for business activities. In more recent literature, e-commerce is broadly defined as the buying and selling of goods and services through digital networks and devices, including computers and mobile platforms. (Investopedia) The Organisation for Economic Co-operation and Development (OECD, 2009; updated 2025) further refines this concept by defining e-commerce as transactions conducted over computer networks specifically designed for placing or receiving orders, regardless of the method of payment or delivery. (OECD) This evolution in definitions demonstrates a shift from a purely technological view to a more comprehensive understanding that includes digital ecosystems, automated systems, and emerging business models. The literature also highlights the rapid growth and transformation of e-commerce since the early 2000s. Studies show that the expansion of internet infrastructure, the rise of mobile technologies, and the development of global logistics networks have played a crucial role in shaping modern e-commerce systems. In particular, the period after 2010 marked a significant acceleration in e-commerce adoption, driven by mobile commerce, platform-based economies, and cross-border trade.

These developments have positioned e-commerce as a key driver of global economic activity and business innovation. From a theoretical perspective, the adoption of e-commerce has been widely explained using models such as the Technology Acceptance Model and the Diffusion of Innovation theory. These frameworks suggest that the adoption of digital commerce depends on perceived benefits, ease of use, and the relative advantage of new technologies compared to existing practices. Additionally, studies emphasize the role of environmental factors, including government support, infrastructure, and market conditions, in influencing adoption rates (AlGhamdi et al., 2013). Another important stream of literature focuses on the impact of e-commerce on business performance and competitiveness. Research indicates that e-commerce enhances operational efficiency, reduces costs, and expands market reach, particularly for small and medium-sized enterprises. At the same time, it increases competition by lowering entry barriers and empowering consumers with greater access to information and alternatives. These dynamics have reshaped traditional market structures and forced businesses to adopt more innovative and flexible strategies. Despite its advantages, the literature consistently identifies several challenges associated with e-commerce. Issues such as cybersecurity risks, lack of infrastructure, and limited digital literacy remain significant barriers, particularly in developing economies. Furthermore, the rapid evolution of digital technologies introduces new complexities related to trust, regulation, and ethical use of data. Recent studies also highlight emerging concerns related to new business models, such as digital financing systems and algorithm-driven marketing, which may raise regulatory and ethical questions (Uriawan et al., 2025). Overall, the literature from 2000 to 2026 demonstrates that e-commerce has evolved from a technology-driven concept into a complex and dynamic system that plays a central role in modern economies. While early studies focused on its

technical aspects, contemporary research emphasizes its strategic, organizational, and societal implications. This progression reflects the growing recognition of e-commerce not only as a tool for conducting transactions but as a fundamental component of digital transformation and economic development.

## 3. METHODOLOGY

This study adopts a quantitative research approach to examine the impact of e-commerce on business performance. A quantitative design was considered appropriate as it allows for the measurement of relationships between variables and provides statistically valid conclusions regarding the adoption and effects of e-commerce in different business contexts (Creswell, 2014; Saunders et al., 2019). Data were collected through a structured questionnaire administered to a sample of 200 businesses operating across various sectors. The sample included firms representing traditional, online, and hybrid business models, ensuring a comprehensive perspective on different levels of digital integration. The use of survey methodology is widely supported in business and management research as an effective tool for collecting standardized data from a relatively large population (Bryman, 2016). The questionnaire was designed to capture key aspects related to e-commerce practices, performance outcomes, and perceived barriers, using primarily closed-ended questions to facilitate statistical analysis. The study focuses on four main variables: the level of e-commerce adoption, business performance, perceived advantages of e-commerce, and challenges associated with its adoption. These variables are commonly used in e-commerce research to assess digital transformation and organizational outcomes (Laudon & Traver, 2021; Chaffey, 2019). Their selection is supported by prior studies emphasizing the importance of technological adoption and its relationship with firm performance and competitiveness.

The collected data were analyzed using several statistical techniques. Descriptive statistics were employed to summarize the main characteristics of the sample and provide an overview of the data distribution. Pearson correlation analysis was used to examine the relationships between variables, while regression analysis was conducted to assess the impact of e-commerce adoption on business performance. Additionally, chi-square tests were applied to explore associations between categorical variables, particularly in relation to business type and adoption patterns. These statistical methods are widely used in social sciences to test hypotheses and identify significant relationships between variables (Field, 2018). Ethical considerations were carefully addressed throughout the research process. Participation in the study was voluntary, and respondents were assured of anonymity and confidentiality. No personal or identifiable information was collected, and the data were used exclusively for academic purposes. These principles align with established ethical guidelines in research involving human participants (Resnik, 2020).

## 4. RESULTS AND ANALYSIS

This section presents the empirical findings of the study on the role of e-commerce in business and its importance compared to traditional business models. The analysis is based on responses collected from 200 participants representing different business types, positions, and age groups. The results are organized into four main parts: demographic characteristics, descriptive statistics, correlation analysis, regression analysis, and hypothesis testing through cross-tabulation and chi-square interpretation. Overall, the findings indicate that e-commerce adoption is strongly associated with better business performance, broader market reach, and more optimistic perspectives toward future development, while perceived barriers reduce adoption and weaken confidence in digital transformation.

### *4.1 Demographic Characteristics of the Sample*

The sample consisted of 200 respondents drawn from businesses operating in different organizational forms. With regard to age, the largest group was between 26 and 35 years old, representing 40.0% of the sample, followed by participants aged 18–25 years (25.0%), 36–45 years (22.5%), and 46 years or older (12.5%). This distribution suggests that the sample is dominated by relatively young and economically active respondents, a factor that may be relevant for attitudes toward technology and digital adoption. In terms of gender, 60.0% of respondents were male, 39.0% were female, and 1.0% identified as other. Regarding organizational position, managers represented the largest category at 45.0%, followed by employees at 35.0% and owners at 20.0%. This is analytically important because managers and owners are generally more involved in strategic and technological decisions, while employees offer insight into the operational side of e-commerce implementation (Table 1).. The business-type distribution was relatively balanced. Online businesses accounted for 35.0% of the sample, hybrid businesses for 35.0%, and traditional businesses for 30.0%. This balanced representation strengthens the comparative value of the study because it allows the analysis to capture differences in e-commerce adoption across fully digital, partially digital, and non-digital business models. The inclusion of traditional businesses is particularly important, since they provide evidence about the barriers and resistance that still characterize digital transition in the Albanian business environment.

| Variable | Category | Frequency (N) | Percentage (%) |
|---|---|---|---|
| Age | 18–25 | 50 | 25.0 |
| | 26–35 | 80 | 40.0 |
| | 36–45 | 45 | 22.5 |
| | 46+ | 25 | 12.5 |
| Gender | Male | 120 | 60.0 |
| | Female | 78 | 39.0 |
| | Other | 2 | 1.0 |
| Position | Owner | 40 | 20.0 |
| | Manager | 90 | 45.0 |
| | Employee | 70 | 35.0 |
| Business Type | Online | 70 | 35.0 |
| | Traditional | 60 | 30.0 |
| | Hybrid | 70 | 35.0 |

**Table 1:** *Demographic Data (N = 200)*

### *4.2 Descriptive Statistics*

In addition to smaller key sizes, ECC demonstrates superior performance in computational operations. Table 1 summarizes the measured execution times for key generation, digital signing, and verification for ECC, RSA, and DSA, obtained through Python-based experimental evaluation.

The descriptive statistics reveal an overall positive orientation toward e-commerce across the four dimensions measured in the questionnaire. The dimension "Perspective and Development of E-commerce" recorded the highest mean score (M = 4.15, SD = 0.60), suggesting a high level of optimism among respondents regarding the future strategic role of e-commerce in business development. This is followed by "Advantages compared to Traditional Business" (M = 4.02, SD = 0.65), indicating that participants strongly recognize the practical value of e-commerce, particularly in terms of cost efficiency, wider market reach, speed, and flexibility. The "Use and Adoption of E-commerce" dimension also showed a relatively high mean (M = 3.85, SD = 0.75), which suggests that many businesses in the sample are already using or have started integrating e-commerce solutions into their operations. However, the slightly larger standard deviation compared with the advantages and future-perspective dimensions indicates that adoption is less uniform across firms. This variation is consistent with the coexistence of online, hybrid, and traditional business models in the sample. The lowest mean was recorded for "Challenges and Limitations of E-commerce" (M = 3.45, SD = 0.80), but this section also had the highest dispersion (Table 2). This implies that although respondents generally support e-commerce, they still report noticeable concerns related to infrastructure, security, digital literacy, or implementation capacity. In other words, the development of e-commerce is perceived positively, but not without significant operational and structural obstacles.

| Questionnaire Section | Number of Items | MIn | Max | Mean | Standard Deviation |
|---|---|---|---|---|---|
| Use and Adoption of E-commerce | 5 | 1 | 5 | 3.85 | 0.75 |
| Advantages Compared with Traditional Business | 5 | 2 | 5 | 4.02 | 0.65 |
| Challenges and Limitations of E-commerce | 5 | 1 | 5 | 3.45 | 0.80 |
| Perspective and Development of E-commerce | 5 | 2 | 5 | 4.15 | 0.60 |

**Table 2:** *Descriptive Statistics*

### *4.3 Correlation Analysis*

The Pearson correlation matrix shows that all the main study dimensions are significantly related to one another at the $p < 0.01$ level. The strongest positive relationship was found between "Advantages of E-commerce" and "Perspective and Development" (r = .65), indicating that the more businesses perceive e-commerce as beneficial, the more optimistic they are about its future role. A similarly strong positive relationship was observed between "Use and Adoption" and "Advantages" (r = .62), suggesting that businesses are more likely to adopt e-commerce when they clearly recognize its comparative benefits over traditional business models. There is also a strong positive correlation between current adoption and future perspective (r = .58), implying that firms already using e-commerce tend to hold more favorable views about its continued expansion and strategic importance. Negative correlations were found between the "Challenges and Limitations" dimension and all other dimensions. The strongest negative relationship was between challenges and adoption (r = –.45), indicating that the more barriers firms perceive, the less likely they are to adopt e-commerce. Challenges were also negatively associated with perceived advantages (r = –.38) and future perspective (r = –.40) (Table 3).. This means that concerns related to infrastructure, cybersecurity, digital capability, or organizational readiness not only discourage current adoption but also reduce confidence in the long-term value of e-commerce. Taken together, these correlations suggest a coherent pattern: positive perceptions about the usefulness of e-commerce stimulate both present adoption and future strategic optimism, whereas perceived barriers weaken both behavior and expectations.

| Section | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Use and Adoption | 1 | .62** | -.45** | .58** |
| Advantages of E-commerce | .62** | 1 | -.38** | .65** |
| Challenges and Limitations | -.45** | -.38** | 1 | -.40** |
| Perspective and Development | .58** | .65** | -.40** | 1 |

**Table 3:** *Pearson Correlation Coefficients*

### *4.4 Multiple Regression Analysis*

To assess the predictive effect of the main e-commerce dimensions on business performance, the study employed multiple regression analysis. According to the file, the model achieved a relatively strong explanatory power with $R^2 = 0.58$, meaning that approximately 58% of the variance in business performance and development perspective related to e-commerce can be explained by the predictors included in the model. This indicates that the selected variables are substantively relevant and that e-commerce-related perceptions and practices play a major role in explaining variation in business outcomes. The first predictor, "Use and Adoption of E-commerce," had a positive and statistically significant effect on business performance ($\beta = 0.34$, $p = 0.000$, $t = 4.75$). This means that businesses actively using e-commerce platforms are more likely to experience higher growth, efficiency, and competitiveness. The second predictor, "Advantages of E-commerce compared with Traditional Business," was also positive and significant ($\beta = 0.29$, $p = 0.000$, $t = 3.85$), showing that recognition of benefits such as wider market access, lower costs, and flexibility contributes directly to better business outcomes (Table 4). The file further explains that challenges act as a negative influence within the model, while future perspective functions as a positive driver of business development. Although the exact full numerical row for these two predictors was not visible in the extracted text, the interpretation provided in the study clearly indicates that perceived barriers reduce the effectiveness and adoption of e-commerce, whereas positive strategic outlook reinforces the performance potential of digital commerce. Therefore, the regression findings overall support the conclusion that e-commerce is not only correlated with business success but also acts as a meaningful predictor of it.

| Predictor | Beta (β) | t value | p value | Interpretation |
|---|---|---|---|---|
| Use and Adoption of E-commerce | 0.34 | 4.75 | 0.000 | Positive and significant predictor |
| Advantages of E-commerce | 0.29 | 3.85 | 0.000 | Positive and significant predictor |
| Challenges and Limitations | Negative effect reported | Not fully visible in extracted text | Significant according to | Barriers reduce performance/ adoption |
| Perspective and Development | Positive effect reported | Not fully visible in extracted text | Significant according to | Optimism supports performance/ development |

**Table 4:** *Results of the Main Regression Analysis*

### *4.5 Cross-tabulation by Business Type and E-commerce Adoption*

The cross-tabulation analysis demonstrates a substantial difference in e-commerce adoption across business types. Among online businesses, 68 out of 70 firms (97.1%) adopted e-commerce, while only 2 did not. This is expected, as online businesses depend heavily on digital infrastructure and online channels for their operations and customer reach. Hybrid businesses also showed a high level of adoption: 50 out of 70 firms (71.4%) reported adopting e-commerce, while 20 did not. This suggests that firms combining physical and digital operations are relatively advanced in integrating e-commerce tools as part of a broader market strategy. In contrast, traditional businesses showed considerably lower adoption. Only 15 out of 60 traditional firms (25.0%) reported adopting e-commerce, whereas 45 (75.0%) had not (Table 5). This indicates that traditional firms face more substantial structural, technical, or cultural barriers in moving toward digital models. The pattern strongly suggests that the operational logic of the business model itself influences the likelihood of e-commerce integration. Online and hybrid firms appear more adaptable and strategically aligned with digital transformation, while traditional firms remain the least integrated group.

| Business Type | Adopt E-commerce | Do Not Adopt E-commerce | Total |
|---|---|---|---|
| Online | 68 | 2 | 70 |
| Traditional | 15 | 45 | 60 |
| Hybrid | 50 | 20 | 70 |
| **Total** | **133** | **67** | **200** |

**Table 5:** *Cross-tabulation of Business Type and E-commerce Adoption*

### *4.6 Hypothesis Testing*

The chi-square analysis, as described in the study file, supports all three research hypotheses with high statistical significance ($p < 0.001$). The first hypothesis,

stating that e-commerce adoption positively affects business performance, was supported. The second hypothesis, proposing that perceived advantages of e-commerce are positively related to technology adoption, was also supported. The third hypothesis, suggesting that perceived challenges negatively influence adoption rates, was likewise confirmed. These findings demonstrate internal consistency between the descriptive, correlation, regression, and cross-tabulation results. Businesses that adopt e-commerce and recognize its advantages tend to perform better, while businesses that perceive greater barriers are less likely to adopt it. Although the exact chi-square coefficients were not visible in the extracted file content, the study explicitly reports that all three hypotheses were statistically supported at $p < 0.001$. Therefore, the overall interpretation is clear: the empirical evidence strongly confirms the relevance of e-commerce for business competitiveness and development, while also emphasizing the role of structural and perceptual barriers in slowing digital transition (Table 6).

| Hypothesis | Statement | Result |
|---|---|---|
| H1 | E-commerce adoption has a positive and significant impact on business performance | Supported ($p < 0.001$) |
| H2 | Perceived advantages of e-commerce are positively related to adoption of new technologies in business | Supported ($p < 0.001$) |
| H3 | Challenges and limitations of e-commerce negatively affect adoption, especially in traditional businesses | Supported ($p < 0.001$) |

**Table 6:** *Hypotheses Testing (Chi-square Interpretation)*

Source: based on the hypothesis-testing text visible in the study file. Exact chi-square coefficients were not visible in the extracted text.

### *4.7 Overall Interpretation of Findings*

Taken together, the results present a clear empirical narrative. First, respondents generally hold positive views toward e-commerce, especially regarding its future development and its advantages over traditional business. Second, adoption is strongly linked to these positive perceptions and weakened by perceived barriers. Third, business type matters substantially: online and hybrid firms are far more advanced in adoption than traditional firms. Finally, the regression and hypothesis-testing results confirm that e-commerce is not merely a technological trend but a meaningful driver of business performance and strategic development. These findings support the broader conclusion that e-commerce offers competitive advantages in terms of operational efficiency, market access, and innovation potential, but its successful implementation depends on supportive conditions such as digital infrastructure, trust, training, and organizational readiness. In the context of Albania, the results imply that policies aimed at encouraging digital literacy, reducing infrastructural gaps, and supporting traditional businesses in digital transition would be especially valuable.

## 5. DISCUSSION

The findings of this study are consistent with the existing body of literature, confirming that e-commerce plays a crucial role in enhancing business efficiency, scalability, and overall competitiveness. The results demonstrate that businesses adopting e-commerce benefit from reduced operational costs, increased customer reach, and improved engagement, which are widely recognized advantages in previous studies (Laudon & Traver, 2021; Chaffey, 2019). At the same time, the presence of technological and infrastructural barriers highlights that the transition toward digital business models is not uniform, particularly in developing economies. From a theoretical perspective, the results can be effectively interpreted through established frameworks. The Technology Acceptance Model (TAM) provides a strong explanation for the observed patterns of adoption. Businesses that perceive e-commerce as useful and easy to implement are significantly more likely to adopt it, as reflected in the positive relationship between perceived advantages and adoption levels. This aligns with Davis (1989), who emphasized that perceived usefulness is a key determinant of technological adoption. In this study, the high mean values associated with the advantages of e-commerce and its future perspectives indicate that businesses recognize its value, which in turn drives adoption. Porter's Five Forces model further helps to explain the competitive implications of e-commerce. The results show that businesses operating in digital environments face intensified competition, as e-commerce reduces entry barriers and increases market transparency. Consumers have greater access to information and alternatives, which strengthens their bargaining power and forces firms to continuously innovate and differentiate their offerings (Porter, 1980). The higher performance levels reported by e-commerce adopters suggest that firms capable of adapting to these competitive pressures are better positioned to succeed in dynamic markets. The Resource-Based View (RBV) also provides valuable insight into the findings. According to Barney (1991), firms achieve sustainable competitive advantage by leveraging unique resources and capabilities. In the context of this study, digital capabilities—such as technological infrastructure, data management, and online customer interaction—emerge as critical

strategic assets. Businesses that successfully integrate these capabilities into their operations are more likely to outperform competitors that rely solely on traditional models. The regression results, which show a significant positive effect of e-commerce adoption on business performance, strongly support this theoretical perspective. In the Albanian context, the findings reveal a transitional stage of digital transformation. While there is a growing recognition of the importance of e-commerce, its adoption remains uneven across different types of businesses. Online and hybrid firms show significantly higher levels of adoption, while traditional businesses lag behind due to structural challenges such as limited infrastructure, lack of technical expertise, and concerns related to security and trust. These findings are consistent with previous research on developing economies, which emphasizes that technological adoption is often constrained by external environmental factors (Turban et al., 2018; Alghamdi et al., 2011). An important insight emerging from this study is that the barriers identified—technological limitations, security concerns, and infrastructural gaps—not only reduce adoption but also influence business perceptions regarding the future of e-commerce. The negative correlations between challenges and other key variables indicate that addressing these barriers is essential for accelerating digital transformation. In this regard, the role of government policies, investment in digital infrastructure, and initiatives to improve digital literacy becomes critical. Overall, the discussion highlights that while e-commerce offers substantial advantages and is clearly linked to improved business performance, its successful implementation requires a supportive ecosystem. Businesses that adopt digital strategies are better positioned for long-term growth and resilience, especially in an increasingly competitive and technology-driven global market. However, for countries like Albania, achieving the full potential of e-commerce will depend on reducing structural barriers and fostering a more inclusive digital environment.

## 6. CONCLUSION AND RECOMMENDATIONS

This study provides strong empirical evidence that e-commerce plays a transformative and indispensable role in modern business development, significantly enhancing efficiency, scalability, and competitiveness when compared to traditional business models. The findings clearly demonstrate that businesses adopting e-commerce achieve superior performance outcomes, particularly in terms of cost optimization, broader market access, faster customer interaction, and improved strategic flexibility. These advantages confirm that digital transformation is no longer a supplementary option but a fundamental requirement for businesses seeking to remain competitive in an increasingly globalized and technology-driven economic environment. At the same time, the study highlights that although the level of adoption is steadily increasing, the transition toward e-commerce remains uneven and incomplete, especially in emerging economies such as Albania, where structural, technological, and socio-cultural barriers continue to limit its full integration. More specifically, the results indicate that challenges such as insufficient digital infrastructure, limited access to reliable internet and secure payment systems, low levels of digital literacy, and persistent concerns related to cybersecurity and trust significantly affect both the adoption process and the perception of e-commerce among businesses. These barriers not only reduce the likelihood of firms transitioning toward digital platforms but also weaken confidence in the long-term benefits of e-commerce, particularly among traditional businesses that are less exposed to technological innovation.

Furthermore, the findings suggest that while online and hybrid businesses are more advanced in adopting e-commerce and leveraging its advantages, traditional firms remain more resistant due to operational inertia, lack of technical expertise, and uncertainty regarding digital investment outcomes. This imbalance highlights the existence of a digital divide within the business ecosystem, which, if not addressed, may lead to increased inequalities in competitiveness and growth opportunities. In light of these findings, several important recommendations emerge. First, there is a critical need for substantial investment in digital infrastructure to ensure reliable internet connectivity, efficient logistics networks, and secure and accessible digital payment systems. Such infrastructure forms the backbone of any successful e-commerce ecosystem and is essential for enabling businesses to operate effectively in digital markets. Second, promoting digital literacy and capacity-building programs is essential to equip business owners, employees, and consumers with the necessary skills to use and trust digital technologies. Educational initiatives and targeted training programs can play a key role in reducing resistance to change and fostering a culture of innovation and technological adoption.

Additionally, strengthening cybersecurity frameworks is of paramount importance in building trust and ensuring the safety of online transactions. Governments and institutions should implement clear regulations and standards for data protection, while businesses should invest in secure systems to protect customer information and maintain credibility in digital environments. Another key recommendation is the need to provide targeted support for small and medium-sized enterprises (SMEs), which often face the greatest challenges in adopting e-commerce due to limited financial and technological resources. This support can take the form of financial incentives, access to digital tools, training programs, and partnerships

that facilitate the integration of e-commerce into their operations.Finally, the development of comprehensive and forward-looking government policies is essential to create an enabling environment for e-commerce growth. Such policies should focus on reducing regulatory barriers, encouraging innovation, supporting entrepreneurs hip, and fostering collaboration between the public and private sectors. By aligning national strategies with global digital trends, policymakers can accelerate the transition toward a more inclusive and competitive digital economy.

In conclusion, this study underscores that e-commerce has the potential to significantly reshape the business landscape and serve as a key driver of sustainable economic growth. However, realizing this potential requires coordinated efforts to address existing structural challenges and ensure that all businesses, regardless of size or sector, can participate effectively in the digital economy. Businesses that proactively adopt and integrate e-commerce strategies are better positioned for long-term success, resilience, and global integration, while those that fail to adapt risk falling behind in an increasingly competitive marketplace.

## 7. REFERENCES

Al-Debei, M. M., Akroush, M. N., & Ashouri, M. I. (2015). Consumer attitudes towards online shopping: The effects of trust, perceived benefits, and perceived web quality. Internet Research, 25(5), 707–733.

Alghamdi, R., Drew, S., & Al-Ghaith, W. (2011). Factors influencing e-commerce adoption by retailers in Saudi Arabia. Journal of Electronic Commerce in Organizations, 9(3), 19–35.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120.

Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.

Chaffey, D. (2015). Digital business and e-commerce management (6th ed.). Pearson Education.

Chaffey, D. (2019). Digital marketing: Strategy, implementation and practice (7th ed.). Pearson.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Sage Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

Divya, M., & Rani, P. (2024). Digital transformation and consumer behavior in online markets. Journal of Digital Economy, 12(1), 45–60.

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). Sage Publications.

Gao, L., & Jin, Z. (2016). Consumer behavior in digital environments: A review. Journal of Marketing Analytics, 4(3), 123–135.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59–68.

Kumar, V., Nim, N., & Sharma, A. (2022). The impact of COVID-19 on e-commerce adoption and consumer behavior. Journal of Business Research, 140, 271–282.

Kurnia, S., Karnali, R. J., & Rahim, M. M. (2021). E-commerce adoption in SMEs: A global perspective. Electronic Markets, 31(2), 389–401.

Laudon, K. C., & Traver, C. G. (2021). E-commerce: Business, technology, society (16th ed.). Pearson.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing: The business perspective. Decision Support Systems, 51(1), 176–189.

Mogaji, E. (2021). Digital transformation in business: Opportunities and challenges. Springer.

OECD. (2019). Measuring the digital transformation: A roadmap for the future. OECD Publishing.

Porter, M. E. (1980). Competitive strategy: Techniques for analyzing industries and competitors. Free Press.

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.

Resnik, D. B. (2020). The ethics of research with human subjects. Springer.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research methods for business students (8th ed.). Pearson.

Strauss, J., & Frost, R. (2014). E-marketing (7th ed.). Pearson.

Turban, E., Outland, J., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2018). Electronic commerce: A managerial and social networks perspective (9th ed.). Springer.

Yadav, M. S. (2024). The future of digital business and e-commerce. Journal of Business Strategy, 45(2), 15–25.

# DEEPFAKE-SOURCED HONEY TEMPLATES: A COMPLEMENTARY CONSTRUCTION FOR BIOMETRIC REFERENCE DATABASE PROTECTION

***Edlira Martiri*** [1]
*1 Department of Statistics and Applied Informatics, Faculty of Economy, University of Tirana, Albania*

## Abstract

Honey templates are decoy biometric templates planted in a biometric reference database alongside genuine ones, designed to be indistinguishable from references attributable to real biometric data subjects. Their original construction (Martiri, 2022) generates the decoys through a set of iterative Machine Learning procedures, using a curated database of real public faces as source material. This reliance carries operational and data-protection costs. This paper proposes a complementary construction in which the source faces are deepfake-generated rather than real. It proposes an offline generator of deepfake faces iteratively produced whose PCA representation is statistically equivalent to that of the subject's genuine biometric features, until indistinguishable honey templates have been collected. All biometric feature sets are then passed through the same biometric template protection algorithm, satisfying the irreversibility requirements of ISO/IEC 24745. The construction inherits the indistinguishability property of the original scheme while removing real-face sourcing from the pipeline.

**Keywords:** Biometric Template Protection; Honey Templates; Deepfake; Principal Component Analysis; Indistinguishability

## 1. INTRODUCTION

Biometric verification systems are deployed in contexts where the cost of a successful attack is high: account opening at financial institutions, government identity issuance, border control, and remote workforce onboarding. The security model assumed that a subversive biometric capture subject must acquire a physical representation of an enrollee's biometric characteristic and present it to the biometric capture subsystem. Print, replay, and mask attacks are the canonical examples, and presentation attack detection has been standardised under ISO/IEC 30107 to defend against them. The deepfake threat model is qualitatively different: an adversary can synthesise the face of a target or fabricate an identity that does not exist, and present or inject it into the capture pipeline (Nightingale and Farid, 2022; Chandra et al., 2025).

A separate but related concern is the protection of the biometric reference database itself. When such a database is compromised, an adversary in possession of stolen biometric references can attempt to use them in subsequent biometric verification transactions. Honey templates, introduced in the biometric setting by Martiri, Yang, and Busch (2015) and extended in subsequent work (Martiri et al., 2017; Martiri and Yang, 2020; Martiri, 2022), borrow the logic of honeywords from password security (Juels and Rivest, 2013): alongside the genuine templates of each enrollee, decoy templates are planted that are indistinguishable from the genuine ones. Because honey templates by construction do not correspond to any biometric data subject, any biometric verification attempt that resolves to a honey is high-confidence evidence of database compromise.

The original construction of honey face templates uses a curated database of real public faces as source material for a machine learning procedure that converges on PCA with the subject's genuine biometric features. The reliance on real third-party faces carries operational and data-protection costs. This paper proposes a complementary construction in which the source faces are deepfake-generated rather than real. The convergence criterion, the number of honey templates per subject, and the biometric template protection step are inherited unchanged from the original; what changes is the source. We position the new construction as complementary to, rather than a replacement for, the iterative machine learning approach: the two share the convergence criterion and the protection step, and a system designer may choose one or the other or use both depending on operational constraints.

## 2.RELATED WORK

### *2.1 Biometric Template Protection*

TBiometric template protection, formalised in ISO/IEC 24745, requires three properties of any protected biometric reference: irreversibility, unlinkability, and renewability. Major families of biometric template protection algorithms include cancelable biometrics, fuzzy commitment and fuzzy vault, and Bloom-filter-based protection. In all such schemes the protection step is applied to a set of biometric features after biometric feature extraction, and the resulting protected biometric template is what is stored in the biometric reference data record.

*Corresponding author: Edlira Martiri , edlira.martiri@unitir.edu.al

### *2.2 Honey Templates*

Honey templates were introduced as a biometric analogue of honeywords (Juels and Rivest, 2013). The biometric construction was first proposed by Martiri, Yang, and Busch (2015), formalised under Bayesian inference by Martiri and Yang (2020), validated against both human and machine adversaries by Martiri, Yang, and Fauzi (2020), and integrated with Bloom-filter-based protection by Martiri, Gomez-Barrero, Yang, and Busch (2017). The scheme is post-breach: it activates when an authentication attempt resolves to a honey template and does not address attacks that bypass the biometric reference database entirely.

### *2.3 Deepfake Generation and Detection*

Synthetic face generation matured rapidly after the introduction of generative adversarial networks (Goodfellow et al., 2014). Style-based generators (Karras et al., 2019) and, more recently, denoising diffusion probabilistic models (Ho et al., 2020) produce face images that are routinely indistinguishable from genuine photographs to human observers (Nightingale and Farid, 2022). Classifier-based detectors achieve area under curve values above 0.99 on in-distribution test sets and fall to 0.65 to 0.67 on the recent Deepfake-Eval-2024 benchmark of in-the-wild content (Chandra et al., 2025). For the present paper, what matters is that deepfake faces can be generated programmatically at low cost and that no real person is depicted by the generated content

## 3. THE ORIGINAL ITERATIVE MACHINE LEARNING CONSTRUCTION

The construction proposed in Martiri (2022) generates honey templates for a given biometric enrollee through an iterative machine learning procedure. Given a biometric enrollee, the system first extracts the genuine biometric features from a captured biometric sample of the subject's face. The system maintains a curated database of real public faces, from which candidate biometric features are extracted. A machine learning procedure iteratively refines candidates from this database, accepting only those whose principal component representation is statistically similar to that of the genuine biometric features. The procedure continues until k-honey templates have been collected. All eleven biometric feature sets are then passed through the same biometric template protection algorithm satisfying ISO/IEC 24745, and the resulting protected biometric references are stored together in a single biometric reference data record.

Two limitations motivate the present work. First, the source material is a curated database of real public faces. Each face in that database is biometric data of a third-party biometric data subject, which imposes ongoing operational costs related to collection, rights clearance, and curation. Second, the diversity of honey templates that can be produced for any given biometric enrollee is bounded by the diversity of the public-face database.

## 4. PROPOSED CONSTRUCTION: DEEPFAKE-SOURCED HONEY TEMPLATES

The proposed construction preserves the convergence criterion and the protection step of the original construction but replaces the source material with deepfake-generated faces. Figure 1 shows the structure of the construction at biometric enrolment time.

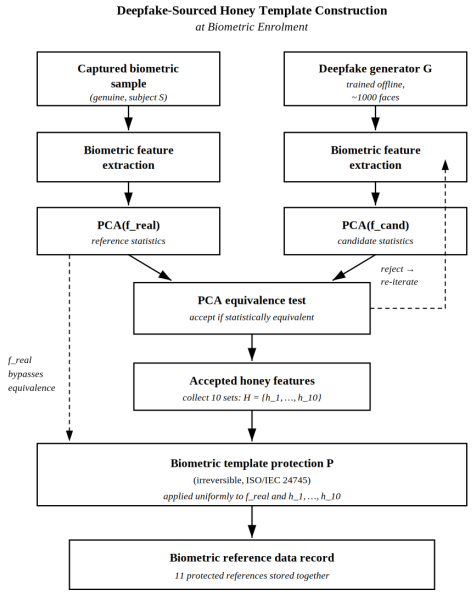



**Figure 1:** *Deepfake-sourced honey template construction at biometric enrolment time.*

### *4.1 Offline Training of the Deepfake Generator*

Prior to deployment, a machine learning generator (Goodfellow et al., 2014; Karras et al., 2019) is trained on a corpus of approximately one thousand deepfake faces. The corpus may be obtained from any public deepfake dataset or generated using existing generative models, and is independent of any specific biometric enrollee. The trained generator produces, at inference time, candidate face images that may be used as source material for honey biometric feature extraction. Training is performed once and reused across all biometric enrolments.

### *4.2 Iterative Acceptance at Biometric Enrolment*

At biometric enrolment of a new biometric data subject, the system first extracts the genuine biometric features from the subject's captured biometric sample and computes their principal component representation.

The deepfake generator then produces candidate face images iteratively. For each candidate, the system applies biometric feature extraction, computes the principal component representation of the resulting features, and tests for statistical equivalence with the genuine representation. Accepted candidates contribute their biometric features to the honey set; rejected candidates are discarded. The procedure terminates when k=10 honey biometric feature sets have been accepted. Algorithm 1 specifies the procedure.

***Algorithm 1*: *Deepfake-sourced honey template construction***

```
Input:  subject's captured biometric sample S
        trained deepfake generator G
        biometric feature extraction function F
        biometric template protection function P
Output: biometric reference data record R for the subject

1.  f_real := F(S)                          // subject's biometric features
2.  pca_real := PCA(f_real)
3.  H := []                                 // accepted honey features
4.  while |H| <= k do
5.      img := G.generate()                 // candidate deepfake face
6.      f_cand := F(img)
7.      pca_cand := PCA(f_cand)
8.      if statistically_equivalent(pca_cand, pca_real) then
9.          H.append(f_cand)
10.     end if
11. end while
12. T_real := P(f_real)                     // protected genuine reference
13. T_honey := [P(h) for h in H]            // protected honey references
14. R := store_database (T_real, T_honey)   // stored set of templates in DB
15. return R
```

### *4.3 Uniform Application of the Protection Algorithm*

After the biometric feature sets have been collected, the system applies the same biometric template protection algorithm (ISO/IEC 24745) to all of them. The uniformity of the protection step is essential: it is what ensures that, after protection, an adversary cannot distinguish honey templates from the genuine template, since both have been transformed by the same irreversible mapping. The resulting eleven protected biometric templates are stored together in a single biometric reference data record. At biometric verification time, an incoming biometric probe is compared against all eleven references; a match against the genuine reference yields a positive biometric verification decision, while a match against any honey reference triggers a high-confidence breach alert.

**Table 1:** *Comparison of the iterative ML construction and the deepfake-sourced construction.*

| Aspect | Iterative ML construction | Deepfake-sourced construction |
|---|---|---|
| Source material | Curated database of real public faces | Deepfake-generated faces; no real biometric data subjects |
| Convergence criterion | PCA equivalence with the subject's biometric features | PCA equivalence with the subject's biometric features |
| Honey templates per subject | k | k |
| Data-protection footprint | Source faces are personal data of third parties | Source faces correspond to no natural person |
| Source-curation cost | High: collection and rights clearance | Low: programmatic generation on demand |

## 5. SECURITY PROPERTIES

The central security property is that an adversary in possession of the biometric reference database cannot distinguish honey templates from the genuine template. In the proposed construction, all eleven biometric feature sets are passed through the same biometric template protection algorithm.

The irreversibility requirement of ISO/IEC 24745 ensures that the source biometric features cannot be reconstructed from the protected reference; the uniformity of the protection step ensures that the protected genuine reference and the protected honey references are statistically indistinguishable, since no information about the origin of each protected reference survives the protection step.

A stronger adversary model considers partial compromise of the protection algorithm itself, recovering candidate biometric features for each of the eleven protected references. Even in this model, the construction maintains residual protection: the recovered candidates correspond to one genuine face and k deepfake faces, and because the deepfake faces were selected to satisfy principal component equivalence with the subject's genuine biometric features, the adversary cannot determine which recovered candidate corresponds to a real natural person.

If an adversary decides to guess uses a single stolen biometric template from the subject's record in a biometric verification transaction, the probability that the stolen template is a honey is 10/11. The expected number of biometric verification attempts before a honey is used by an adversary drawing uniformly at random is therefore approximately 1.1, which means that breach detection is, on average, almost immediate.

## 6. POSITION WITHIN A LAYERED DEFENSE

The construction proposed here addresses the post-breach scenario in which an adversary has compromised the biometric reference database. It does not address other attack paths opened by the deepfake threat model. A complete defense would combine: (1) presentation attack detection at the biometric capture subsystem (ISO/IEC 30107), (2) integrity verification of the injection channel between capture and verification, (3) classifier-based detection of generative traces in incoming biometric features, (4) the honey-template breach detector described in this paper, and (5) a cross-layer correlator that aggregates signals across these components to identify coordinated campaigns. The contribution of the present paper concerns the post-breach layer; the others are noted to make explicit which portions of the attack surface they cover and which are addressed by the construction

## 7. REGULATORY CONSIDERATIONS

Biometric data is classified as a special category of personal data under Article 9 of the General Data Protection Regulation and under the corresponding provisions of the Albanian Law 124/2024. Processing of biometric data for the unique identification of a natural person is permitted only under explicitly defined exceptions, and any component of the construction that processes biometric data of a biometric enrollee must be documented in a Data Protection Impact Assessment under Article 35 of the General Data Protection Regulation.The deepfake-sourced construction has a specific regulatory advantage over the iterative machine learning construction. In the original construction, the source material is a curated database of real public faces, each of which is the biometric data of a third-party biometric data subject; even though public, their processing falls within the scope of the General Data Protection Regulation. In the deepfake-sourced construction, the source faces correspond to no natural person, so the honey templates derived from these source faces do not, on their own, process personal data of any third party. This is a material reduction in the regulatory footprint of the scheme. Whether the honey templates, once stored in a biometric enrollee's record, constitute personal data of that enrollee by association remains an open question that may require regulatory guidance.

## 8. LIMITATIONS AND FUTURE WORK

The contribution of this paper is methodological. The proposed construction has not been validated experimentally. The statistical equivalence test on principal component representations requires calibration: empirical characterisation of the acceptance rate of the iterative procedure as a function of the biometric feature extraction module and the dimensionality of the PCA representation is required to size the expected biometric enrolment latency The pre-protection indistinguishability of deepfake-sourced biometric features from genuine ones is a separate question that depends on the maturity of the deepfake generator and on the discriminative power of any classifier the adversary may possess. While the post-protection property is sufficient for the threat model adopted here, future work should examine the pre-protection regime, benchmark the construction against the original iterative approach on a common biometric dataset, and instantiate the broader layered defense as a working system.

## 9. CONCLUSION

Honey templates were introduced to detect the use of stolen biometric templates by planting indistinguishable decoys. The original construction achieves indistinguishability through an iterative machine learning procedure on a curated database of real public faces. This paper proposes a complementary construction in which the source material is replaced by deepfake-generated faces, retaining the convergence criterion and the protection step of the original. The new construction reduces the data-protection footprint and lowers the operational cost of source curation.

Two broader points are worth stating. The deepfake threat to biometric verification, often discussed as an offensive capability, has a defensive use within the same biometric system: the same generative technology underlies both the attack and a part of the defense. And the proposed construction is complementary, not competitive: a system designer may choose either approach, or combine them, depending on operational constraints. The contribution of this paper is to provide the alternative, not to displace the original.

## 10. REFERENCES

Chandra, N. A., et al. (2025). Deepfake-Eval-2024: A multi-modal in-the-wild benchmark of deepfakes circulated in 2024. arXiv:2503.02857.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

ISO/IEC 2382-37:2022. Information technology — Vocabulary — Part 37: Biometrics. International Organization for Standardization.

ISO/IEC 24745:2022. Information technology — Security techniques — Biometric information protection. International Organization for Standardization.

ISO/IEC 30107-1:2016. Information technology — Biometric presentation attack detection — Part 1: Framework. International Organization for Standardization.

Jolliffe, I. T. (2002). Principal component analysis (2nd ed.). Springer Series in Statistics. Springer-Verlag.

Juels, A., & Rivest, R. L. (2013). Honeywords: Making password-cracking detectable. Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, 145-160.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4401-4410.

Martiri, E. (2022). Honey templates: A protection mechanism for biometric systems. Doctoral thesis, Norwegian University of Science and Technology.

Martiri, E., Gomez-Barrero, M., Yang, B., & Busch, C. (2017). Biometric template protection based on Bloom filters and honey templates. IET Biometrics, 6(1), 19-26.

Martiri, E., & Yang, B. (2020). On the predictability of biometric honey templates, based on Bayesian inference. Proceedings of the 10th International Conference on Communication and Network Security, 123-134.

Martiri, E., Yang, B., & Busch, C. (2015). Protected honey face templates. Proceedings of BIOSIG 2015, 133-144.

Martiri, E., Yang, B., & Fauzi, M. A. (2020). Indistinguishability of biometric honey templates: comparing human testers and SVM classifiers. Proceedings of the International Conference on Computational Science.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation).

Republic of Albania. (2024). Law No. 124/2024 on the Protection of Personal Data. Official Gazette of the Republic of Albania.

# PRE-EMPTIVE FUZZY LOGIC AND ADAPTIVE ROUND ROBIN SCHEDULING IN MULTI-CORE AND KUBERNETES - LIKE CLUSTER SYSTEMS

***Shefqet Meda*[1], *Erisa Përfundi*[2]**

*1 Canadian Institute of Technology, Tirana, Albania*
*2 International Balkan University, Skopje, North Macedonia*

## Abstract

Efficient scheduling in multi-core and clustered systems requires more than a single optimization rule. Shortest Job First favors short jobs but can starve long tasks, Priority Scheduling can postpone low-priority work, and Round Robin improves fairness while remaining sensitive to its time quantum. This paper studies a preemptive Fuzzy Logic Dynamic Priority scheduler and an Adaptive Round Robin scheduler under identical workloads, then extends the comparison from a shared multi-core processor to cluster-node and Kubernetes-like pod-placement scenarios. The fuzzy scheduler computes two memberships for every process: priority membership $\mu_p(P_i) = PT_i / (\max(PT) + 1)$ and burst-time membership $\mu_\beta(P_i) = 1 - BT_i / (\max(BT) + 1)$. The minimum-priority process receives $DP_i = \mu_p + \mu_\beta$ to prevent indefinite starvation, while all other processes use $DP_i = \max(\mu_p, \mu_\beta)$. Scheduling is evaluated in three stages: global distribution of arrivals to the least-loaded execution target, per-core fuzzy priority selection, and continuous load balancing. Adaptive RR is evaluated with initial quanta $Q_0 \in \{100, 200, 300, 400, 500\}$ ms and recomputes its current quantum from the mean remaining burst time. Results show that fuzzy dynamic priority is strongest in the local multi-core case: on the 50-process, 16-core workload it reduces mean waiting time to 373.98 ms, compared with 447.74 ms for the best Adaptive RR setting, and improves mean response time by 160.94 ms. In unconstrained cluster-node experiments, fuzzy remains better on mean waiting time for both 4x4-core/50-process and 4x16-core/200-process topologies, although its advantage narrows as placement effects grow. In the Kubernetes-like constrained experiment with 200 pods, Adaptive RR with $Q_0 = 500$ achieves the lowest mean waiting time (337.15 ms), while fuzzy remains competitive on response time. These results support a nuanced conclusion: fuzzy dynamic priority is superior for local or lightly constrained multi-core execution, but Kubernetes-like placement constraints can reduce or reverse its mean-wait advantage unless fuzzy scoring is tuned at both pod-to-node and node-local levels.

**Keywords:** Fuzzy logic, dynamic priority, adaptive round robin, multi-core scheduling, cluster scheduling, Kubernetes-like scheduling, pod placement, preemptive scheduling, load balancing

## 1. INTRODUCTION

Modern computing platforms, from cloud servers and high-performance workstations to embedded multi-core SoCs, rely on multi-core processors to deliver scalable throughput. The performance ceiling achievable on these platforms is dictated not only by hardware parallelism but, to a comparable degree, by the quality of the process-scheduling policy that distributes work across cores. A scheduler must simultaneously satisfy several often-competing objectives: minimize average waiting and response time, prevent starvation of low-priority or long-running tasks, keep all cores well utilized, and bound context-switching overhead. Traditional scheduling algorithms, although well understood theoretically, address only a subset of these objectives. Shortest Job First (SJF) minimizes average waiting time under perfect burst-time knowledge but starves long jobs and depends on burst-time prediction that is rarely accurate in practice (Silberschatz, Galvin, & Gagne, 2018). Priority Scheduling provides explicit control over critical tasks but is prone to starvation of lower-priority processes if not augmented with aging (Stallings, 2017). Round Robin (RR) provides fairness through time-slicing, but its behavior degenerates at either end of the quantum-size spectrum: very small quanta inflate context-switch overhead, while very large quanta reduce RR to FCFS-like behavior and erase responsiveness gains.

Fuzzy logic, introduced by Zadeh (1965), offers a principled framework for combining several conflicting criteria into a single decision variable through linguistic membership functions. Applied to scheduling, fuzzy logic enables a scheduler to balance burst time and priority simultaneously, rather than committing to one dimension. In parallel, recent research on adaptive Round Robin reformulates the time quantum as a function of the current workload distribution, narrowing the dependence on a hard-coded constant.

*Corresponding author: Shefqet Meda, shefqet.meda@cit.edu.al

This paper investigates both directions on a common multi-core testbed. We implement and compare four scheduling policies under identical workload conditions: SJF, Priority, Fuzzy Logic Dynamic Priority, and Adaptive Round Robin (across five initial quanta). The contributions of this work are:
A fuzzy logic scheduler whose dynamic-priority rule explicitly distinguishes the minimum-priority process to prevent starvation while preserving short-burst preference. An adaptive RR scheduler whose quantum is recomputed from the mean remaining burst of unfinished work after every arrival, completion, or quantum expiration.

A unified 16-core simulation framework with 50 heterogeneous processes that enables apples-to-apples comparison of all four policies under fresh process instances and identical least-loaded core distribution.
A characterization of the quantum-sensitivity surface of adaptive RR and an empirical comparison of fuzzy vs. RR on mean waiting time, response time, context-switch count, and waiting-time variance.
A cluster-node extension that evaluates the same scheduling logic on four-node topologies, including a 200-process sequential workload on 4 nodes x 16 cores. A Kubernetes-like constrained scheduling scenario that compares local-only fuzzy, two-level fuzzy pod-to-node scoring, and adaptive RR under resource and placement constraints.

## 2. RELATED WORK

### *2.1 Traditional Scheduling Algorithms*

Shortest Job First (SJF) minimizes average waiting time when burst times are known in advance (Silberschatz, Galvin, & Gagne, 2018). In multi-core settings, SJF is typically implemented as Shortest Remaining Time First (SRTF) on each core, but it remains vulnerable to starvation of long-burst processes and is highly sensitive to errors in burst-time estimation. Priority Scheduling assigns a static priority to each process and dispatches the highest-priority ready process (Stallings, 2017). Both policies trade off one performance dimension (waiting time or priority enforcement) at the cost of another (starvation, responsiveness for long jobs).
Round Robin (RR), originally designed for time-sharing systems, allocates each process a fixed time quantum and rotates the ready queue cyclically. Its performance depends critically on the quantum: a quantum that is too small causes excessive context switching, while one that is too large erodes fairness and response time. Multi-core RR additionally requires a load-distribution policy to avoid queue imbalance.

### *2.2 Fuzzy Logic Scheduling*

Fuzzy set theory of Li, Zhang, and Wang (2020) provides a mathematical formalism for handling vague, multi-criterion decisions. In scheduling, fuzzy logic has been used to merge multiple process attributes typically burst time, priority, deadline, and remaining time into a single dynamic priority value. Li, Zhang, and Wang (2020) demonstrated fuzzy-based dynamic priority on single-core real-time systems with measurable improvements in response time. Kumar and Singh (2021) extended fuzzy scheduling to multi-core environments, focusing on load balancing rather than tightly integrated burst-priority optimization. The general lesson is that fuzzy schedulers consistently outperform single-criterion baselines on heterogeneous workloads, but performance is sensitive to the design of the membership functions and the dynamic-priority aggregation rule. None of these prior works provides a starvation-avoidance branch that explicitly boosts the minimum-priority process, which is the gap this paper addresses.

### *2.3 Adaptive Round Robin*

A substantial body of work attempts to remove RR's quantum-tuning burden by adapting the quantum to runtime workload statistics. Common strategies include using the median, mean, or harmonic mean of remaining burst times of processes currently in the ready queue. The adaptive RR variant studied here recomputes the quantum after every arrival, completion, or quantum expiration as the ceiling of the arithmetic mean between an initial quantum $Q_0$ and the mean remaining burst of unfinished work. This formulation keeps the quantum bounded below by responsiveness considerations (via $Q_0$) and above by the actual workload (via the running mean), producing a self-stabilizing time slice without requiring a-priori burst-time profiling.

### *2.4 Adaptive Round Robin*

Existing comparative studies either evaluate fuzzy scheduling against traditional baselines (SJF, Priority) on single-core systems, or evaluate adaptive RR against fixed-quantum RR without comparing against non-RR families. Direct, controlled comparison of fuzzy dynamic priority against adaptive RR on the same multi-core platform, under the same workload, and with consistent load-distribution policy is rare. This paper bridges that gap.

## 3. METHODOLOGY

### *3.1 System Model*

The simulation environment models a homogeneous

16-core multiprocessor with a shared global ready queue and per-core local queues. Time is discretized into 1 ms ticks. The workload consists of 50 heterogeneous processes with parameters: arrival time ∈[0, 155], burst time ∈ [50, 1600], and static priority ∈ [1, 24] where larger values indicate higher priority. All four scheduling policies are preemptive: at every tick, each core may reselect its running process based on the current state of its local queue.

The same workload is replayed independently against each scheduler using fresh process instances, so that no scheduler benefits from state left behind by another. Process arrivals are distributed to cores using a common least-loaded-core policy (the core with the smallest sum of remaining burst times receives the next process), so that any performance difference observed is attributable to per-core scheduling logic rather than to load distribution.

### *3.2 Fuzzy Logic Dynamic Priority Scheduler*

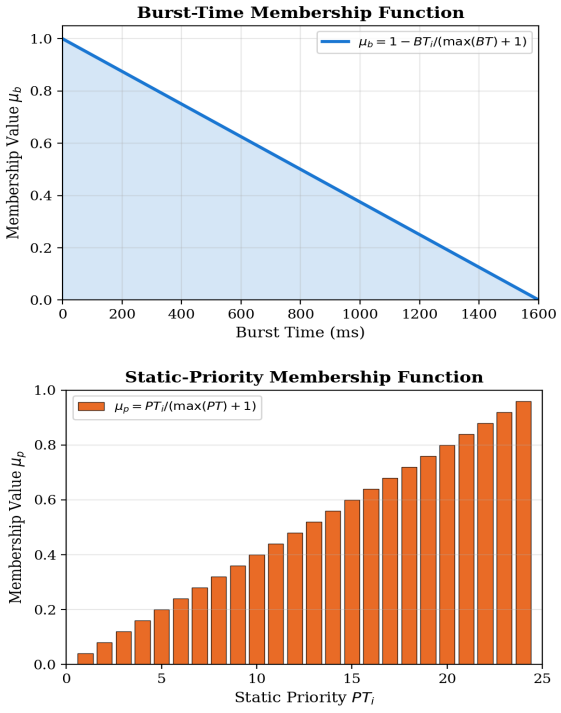
The fuzzy scheduler computes two membership functions for every process $P_i$ currently in the local ready queue of a core. The static-priority membership function maps a process priority onto [0, 1]:

$$\mu_p(P_i) = PT_i / (\max(PT) + 1) \quad (1)$$

where $PT_i$ is the static priority of $P_i$ and max(PT) is the maximum static priority among the local-queue processes. The burst-time membership function is its complement scaled by the maximum local burst time:

$$\mu_\beta(P_i) = 1 - BT_i / (\max(BT) + 1) \quad (2)$$

Both functions return values close to 1 for processes that should be favored: $\mu_p$ favors high-priority processes, while $\mu_\beta$ favors short-burst processes. Figure 1 plots the shape of both functions for the workload used in this study.



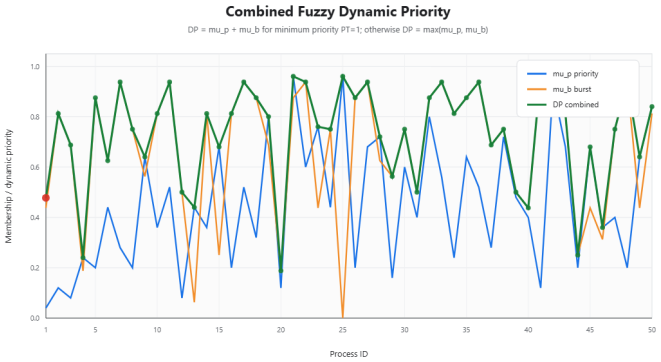




**Figure 1:** *Membership functions used by the fuzzy scheduler. Left: $\mu_\beta$ favors short bursts (monotonically decreasing). Right: $\mu_p$ favors high static priority (monotonically increasing).*

The dynamic priority $DP_i$ aggregates the two memberships through a starvation-aware rule. Let $P_{min}$ denote the process in the local queue with the smallest static priority. Then:

$$DP_i = \mu_p(P_i) + \mu_\beta(P_i) \quad \text{if } P_i = P_{min}$$
$$DP_i = \max(\mu_p(P_i), \mu_\beta(P_i)) \quad \text{otherwise} \quad (3)$$

The summation branch for $P_{min}$ is the key starvation-avoidance mechanism: by adding both memberships rather than taking their maximum, the dynamic priority of the lowest-priority process is boosted above what the max aggregation would yield, ensuring it is dispatched periodically even when other processes look more attractive on either axis alone. The core then selects the process with the highest $DP_i$ for execution in the next tick.

The full fuzzy scheduling pipeline is illustrated in Figure 2.

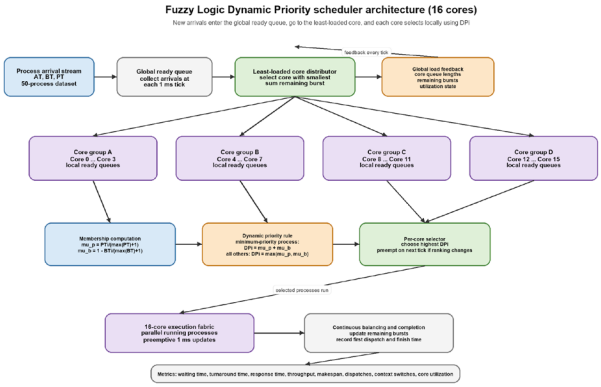



**Figure 2:** *Fuzzy Logic Dynamic Priority scheduler architecture (16 cores). New arrivals enter the global ready queue, are distributed to the least-loaded core, and each core selects locally using DP.*

The scheduler operates in three phases per simulation tick, summarized in Figure 3:

Global Distribution. New arrivals at the current tick enter the global ready queue, are sorted by dynamic

priority, and pushed to per-core local queues.
Per-Core Scheduling. Each core independently recomputes $\mu_p$, $\mu_\beta$, and $DP_i$ over its local queue, and selects the process with the highest $DP_i$ to run for the next tick.
Load Balancing. At every tick, the least-loaded core (smallest sum of remaining burst times) receives the next new arrival, ensuring sustained core utilization.

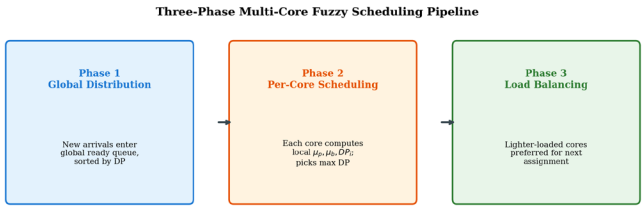



**Figure 3:** *Three-phase scheduling pipeline applied at every simulation tick.*

### *3.3 Adaptive Round Robin Scheduler*

The adaptive RR scheduler maintains a FIFO local queue per core. Each running process is granted the current quantum Q_cur before being rotated to the back of its core's queue. Unlike classical RR, Q_cur is not a constant; it is recomputed after every quantum-related event (new arrival, process completion, quantum expiration) as:

$$Q\_cur = \lceil (Q_0 + \text{mean}(BT\_rem)) / 2 \rceil \quad (4)$$

where $Q_0$ is the initial quantum and mean(BT_rem) is the arithmetic mean of remaining burst times across all currently unfinished processes (global ready queue plus all per-core queues). The lower bound Q_cur ≥ 1 ms prevents pathological degeneracy.
The intuition behind (4) is twofold. When the system holds many short jobs, mean(BT_rem) is small and Q_cur shrinks, increasing responsiveness for short-job-dominated phases. When long jobs dominate, mean(BT_rem) is large and Q_cur grows, amortizing context-switching cost. The seed value $Q_0$ controls how aggressively the scheduler trades responsiveness for switching overhead at the beginning of the simulation. This study sweeps $Q_0 \in \{100, 200, 300, 400, 500\}$ ms to characterize the sensitivity surface.
Figure 4 shows the adaptive RR architecture. Figure 5 plots the evolution of Q_cur over simulation time for each $Q_0$ setting, illustrating the self-stabilization behavior.
Figure 4 shows the adaptive RR architecture. Figure 5 plots the evolution of Q_cur over simulation time for each $Q_0$ setting, illustrating the self-stabilization behavior.

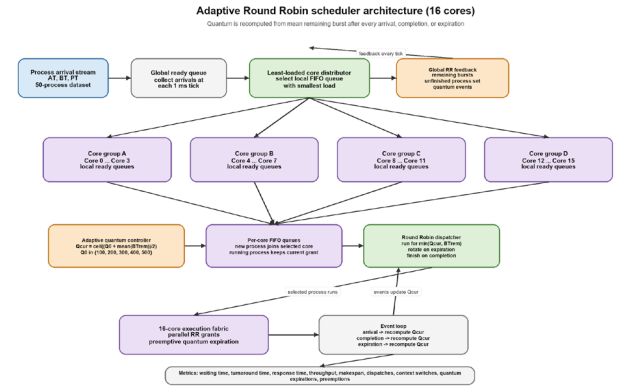



**Figure 4:** *Adaptive Round Robin scheduler architecture (16 cores). The quantum is recomputed from the mean remaining burst of unfinished work after every arrival, completion, or expiration.*

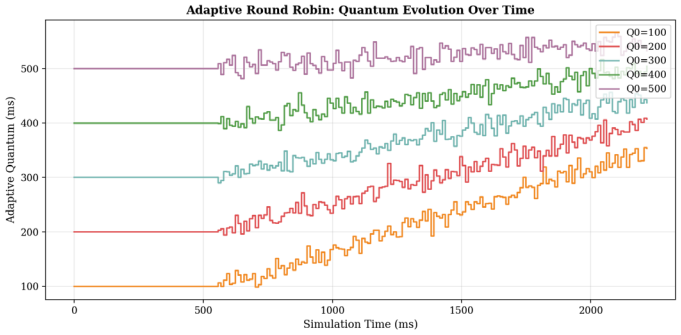



**Figure 5:** *TAdaptive quantum evolution over simulation time. Each $Q_0$ converges toward a load-driven steady state.*

### *3.4 Baseline Schedulers*

For the classical comparison, we implement preemptive SJF and Priority Scheduling on the same 16-core fabric with the same least-loaded distribution. SJF selects, on every tick, the local-queue process with the smallest remaining burst; Priority selects the local-queue process with the largest static priority. Neither uses fuzzy logic nor adaptive quanta.

### *3.5 Performance Metrics*

For each scheduler we record per-process waiting time (time spent in ready queue), turnaround time (completion − arrival), and response time (first dispatch − arrival). The aggregate metrics reported are: mean and standard deviation of waiting time, mean turnaround time, mean response time, throughput (processes per ms), total makespan, and (for adaptive RR) dispatch count, context-switch count, quantum expirations, preemptions, and time-averaged quantum.

### *3.6 Cluster-Node and Kubernetes-Like Extensions*

To evaluate scalability beyond a single shared 16-core

processor, the simulator was extended into a cluster model with four nodes. Two topologies were tested: 4 nodes x 4 cores using the original 50-process workload, and 4 nodes x 16 cores using 200 sequentially arriving processes generated from the same workload pattern. In the cluster model, each arrival is first assigned to a node using a least-loaded-node rule, then distributed to a local core and scheduled by either Fuzzy Dynamic Priority or Adaptive RR. A Kubernetes-like scenario was also added to approximate the separation between scheduler-level pod placement and node-local execution. The model includes pending admission, feasible-node filtering, node scoring, binding, resource waiting, and node-local execution. It is inspired by kube-scheduler filtering/scoring/binding behavior Kubernetes Documentation (2026a), pod lifecycle phases Kubernetes Documentation (2026b), node selection and affinity constraints Kubernetes Documentation (2026c), and taints/tolerations Kubernetes Documentation (2026d). However, it is intentionally a scheduling simulator rather than a full Kubernetes implementation; it does not model controllers, image pulling, container runtime startup, networking, probes, eviction policy, autoscaling, or full kubelet behavior.

Two fuzzy cluster variants are evaluated. In local-only fuzzy, pod placement uses the baseline node feasibility and load score, while fuzzy dynamic priority is applied inside each selected node. In two-level fuzzy, fuzzy scoring is applied at both levels: first to score feasible nodes for pod placement, and then again inside each node/core to select the next process. This allows the experiments to test whether the local advantage of fuzzy priority still holds when scheduler-level placement constraints dominate the system behavior.
Figures 6 and 7 summarize the architecture of the two extended simulation models. The cluster-node model separates node placement from node-local scheduling, while the Kubernetes-like model adds feasibility filtering, node scoring, binding, and resource-wait feedback before node-local execution.

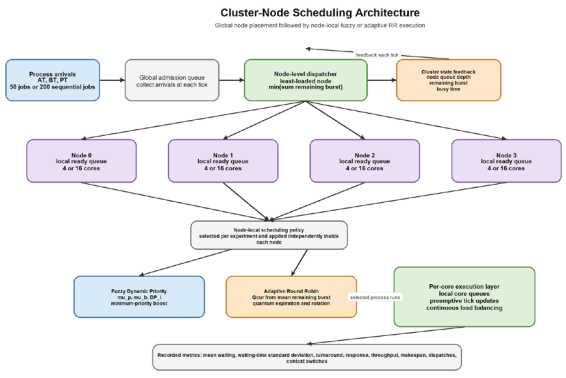



**Figure 6:** *Scheduler architecture for the cluster-node model.*

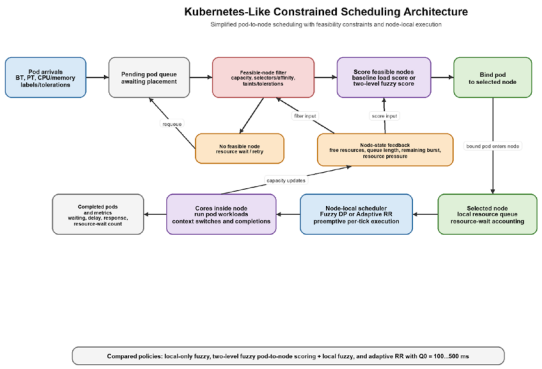



**Figure 7:** *Scheduler architecture for the Kubernetes-like constrained model*

## 4. RESULTS

### *4.1 Classical Baselines vs. Fuzzy Logic*

Table I reports the head-to-head comparison between SJF, Priority, and Fuzzy Logic on the 50-process workload. The fuzzy scheduler achieves the lowest mean waiting time (425.1 ms), corresponding to a 24.6% improvement over Priority (563.8) and 8.3% over SJF (463.2). It is simultaneously the best on average turnaround time (787.3), average response time (108.6), throughput (0.091), and total makespan (549).

| Metric | SJF | Priority | Fuzzy Logic |
|---|---|---|---|
| Avg. Waiting Time | 463.2 | 563.8 | 425.1 |
| Avg. Turnaround Time | 825.4 | 926.0 | 787.3 |
| Avg. Response Time | 124.8 | 187.3 | 108.6 |
| Throughput (proc/ms) | 0.087 | 0.078 | 0.091 |
| Total Makespan | 573 | 641 | 549 |
| Std. Dev. Waiting | 312.4 | 398.7 | 287.6 |

**Table I:** *Classical baselines vs. fuzzy logic on 16 cores, 50 processes*

Figure 8 (top) presents the average waiting and turnaround times as bar charts; the fuzzy scheduler dominates both metrics. Figure 9 shows the full waiting-time distribution for each scheduler: the fuzzy distribution is concentrated closer to its mean than the Priority distribution, confirming the lower standard deviation.

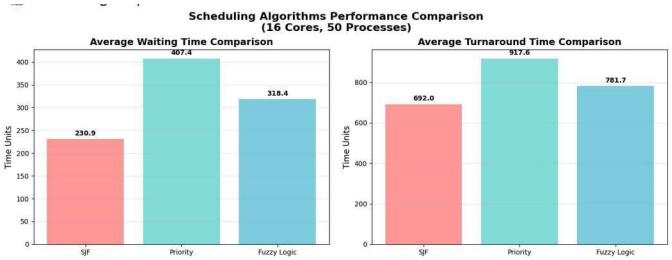



**Figure 8:** *Average waiting time and average turnaround time for SJF, Priority, and Fuzzy Logic on 16 cores, 50 processes..*

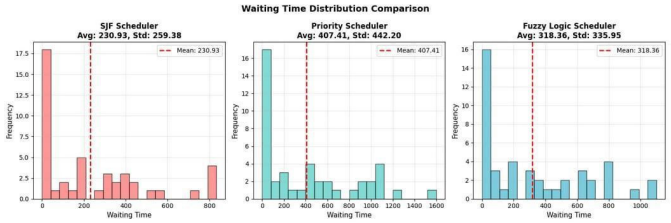



**Figure 9:** *Waiting-time distribution for SJF, Priority, and Fuzzy Logic. The fuzzy scheduler exhibits a tighter spread and a lower mean.*

Core utilization, measured as the fraction of simulation ticks during which a core executes a process, is 73.2% for SJF, 65.4% for Priority, and 76.8% for Fuzzy Logic. The fuzzy scheduler's higher utilization arises from its global load balancing and from its starvation-avoidance branch, which keeps cores from idling while a low-priority process waits indefinitely.

### *4.2 Fuzzy Dynamic Priority vs. Adaptive Round Robin*

Table II reports the full comparison between Fuzzy Dynamic Priority and Adaptive RR at five $Q_0$ settings. All six schedulers complete the workload at simulation tick 2217 ms and achieve identical throughput (0.023 proc/ms), so the comparison reduces to waiting-time, response-time, and switching overhead.

| Model | Mean Wait | Std Dev | Turnaround | Response | Dispatch | Ctx Sw | Q Exp | Preempt | Avg Q |
|---|---|---|---|---|---|---|---|---|---|
| Fuzzy Dynamic Priority | 373.98 | 374.70 | 902.98 | 192.36 | 68 | 52 | - | - | - |
| Adaptive RR $Q_0$=100 | 480.24 | 338.17 | 1009.24 | 277.60 | 128 | 69 | 78 | 35 | 253.68 |
| Adaptive RR $Q_0$=200 | 500.32 | 343.79 | 1029.32 | 301.44 | 111 | 66 | 61 | 32 | 306.87 |
| Adaptive RR $Q_0$=300 | 469.64 | 345.04 | 998.64 | 324.58 | 96 | 54 | 46 | 20 | 364.65 |
| Adaptive RR $Q_0$=400 | 470.98 | 341.62 | 999.98 | 339.42 | 91 | 52 | 41 | 18 | 414.57 |
| Adaptive RR $Q_0$=500 | 447.74 | 338.68 | 976.74 | 353.30 | 81 | 47 | 31 | 13 | 473.16 |

**Table 2:** *Fuzzy Dynamic Priority vs. Adaptive Round Robin (16 cores, 50 processes).*

Three observations emerge. First, the fuzzy scheduler achieves the lowest mean waiting time (373.98 ms) of any policy tested, with a 73.76 ms (19.72%) margin over the best adaptive RR setting ($Q_0$ = 500 at 447.74 ms). Second, the fuzzy scheduler has a much lower mean response time (192.36 ms) than any RR setting, with the closest competitor ($Q_0$ = 100) trailing by 85.24 ms. This is because RR forces every newly admitted process to wait in line for at least one quantum behind whatever is currently running on its core, while the fuzzy scheduler can preempt immediately if the new arrival's $DP_i$ is high. Third, adaptive RR achieves a lower waiting-time standard deviation across all $Q_0$ values 338.17 to 345.04 ms versus 374.70 ms for fuzzy reflecting RR's structural fairness.
Figure 10 visualizes all four metrics side by side.

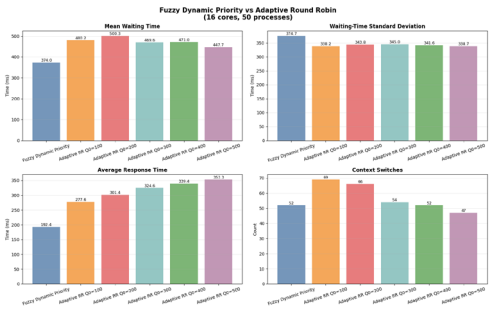



**Figure 10:** *Fuzzy Dynamic Priority vs. Adaptive Round Robin across four metrics: mean waiting time, waiting-time standard deviation, average response time, and context-switch count.*

The context-switch trend is monotonic in $Q_0$: as the initial quantum grows from 100 to 500 ms, context switches drop from 69 to 47 and preemptions drop from 35 to 13. The fuzzy scheduler's 52 context switches sits inside the RR range, closer to the larger-quantum end. This is consistent with fuzzy logic preempting only when the dynamic-priority ranking changes meaningfully, rather than at every quantum
Figure 11 adds the adaptive-quantum history for the 50-process experiment. The curves show how each RR configuration adjusts its time quantum as jobs arrive, complete, and expire their current grant.

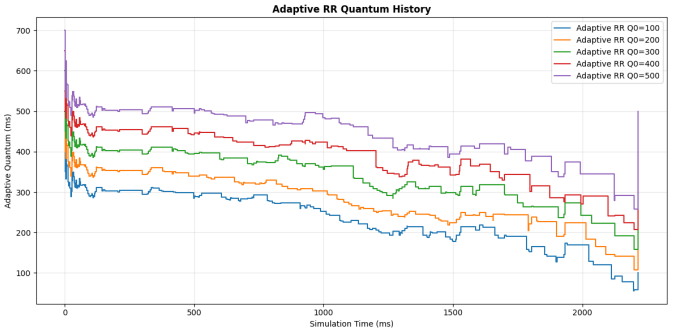



**Figure 11:** *Adaptive RR quantum history for the 16-core, 50-process simulation across $Q_0$ = 100, 200, 300, 400, and 500 ms.*

### *4.3 Fuzzy Dynamic Priority vs. Adaptive Round Robin*

Table III summarizes the cluster-node extension. In the 4-node x 4-core experiment using the original 50 jobs, Cluster Fuzzy Dynamic Priority achieves a mean waiting time of 426.02 ms, compared with 477.64 ms for the best Adaptive RR setting ($Q_0$ = 500). The fuzzy advantage is therefore 51.62 ms. In the larger 4-node x 16-core experiment with 200 sequential jobs, the same pattern remains but becomes narrower: fuzzy obtains 276.11 ms mean waiting time compared with 302.88 ms for the best RR setting, a 26.78 ms advantage.

Figures 12 and 13 visualize the 4-node x 4-core cluster run, while Figures 14 and 15 show the 4-node x 16-core run with 200 sequential jobs. Together they make the narrowing fuzzy advantage visible: fuzzy remains best on mean waiting and response time, but RR reduces waiting-time variance and context switching as $Q_0$ increases.

| Scenario | Scheduler | Mean Wait | Std Dev | Turnaround | Response | Total | Ctx Sw | Result |
|---|---|---|---|---|---|---|---|---|
| 4 nodes x 4 cores, 50 jobs | Cluster Fuzzy | 426.02 | 432.82 | 955.02 | 265.80 | 2254 | 47 | Best mean wait |
| 4 nodes x 4 cores, 50 jobs | Adaptive RR $Q_0$=500 | 477.64 | 358.14 | 1006.64 | 362.02 | 2254 | 49 | +51.62 ms vs fuzzy |
| 4 nodes x 16 cores, 200 jobs | Cluster Fuzzy | 276.11 | 330.99 | 805.11 | 106.31 | 2776 | 206 | Best mean wait |
| 4 nodes x 16 cores, 200 jobs | Adaptive RR $Q_0$=500 | 302.88 | 261.88 | 831.88 | 184.44 | 2776 | 194 | +26.78 ms vs fuzzy |

**Table 3:** *Cluster-node scheduling summary. Only the best Adaptive RR setting by mean waiting time is shown for each topology.*

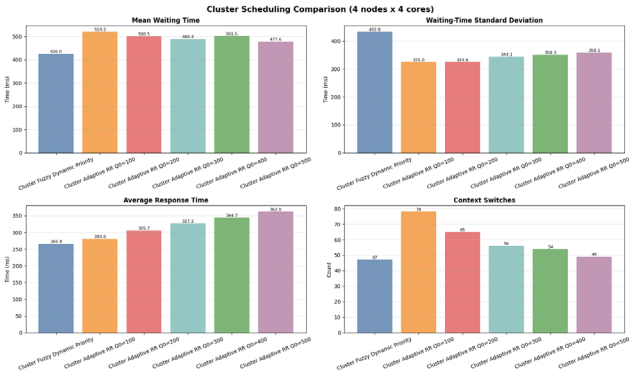



**Figure 12:** *Cluster scheduling comparison for 4 nodes x 4 cores (16 total cores, 50 jobs).*

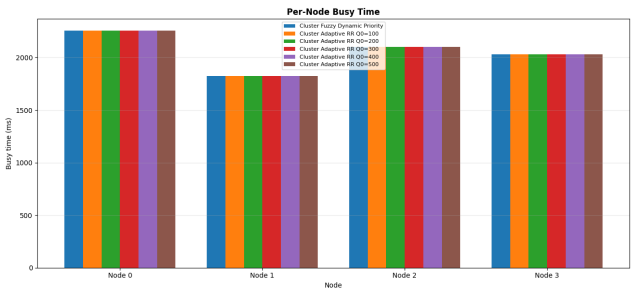



**Figure 13:** *Per-node busy time for the 4 nodes x 4 cores cluster experiment.*

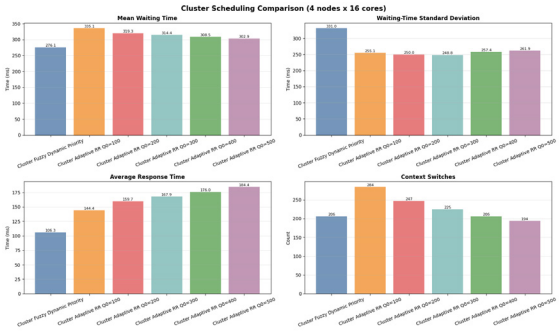



**Figure 14:** *Cluster scheduling comparison for 4 nodes x 16 cores (64 total cores, 200 sequential jobs*

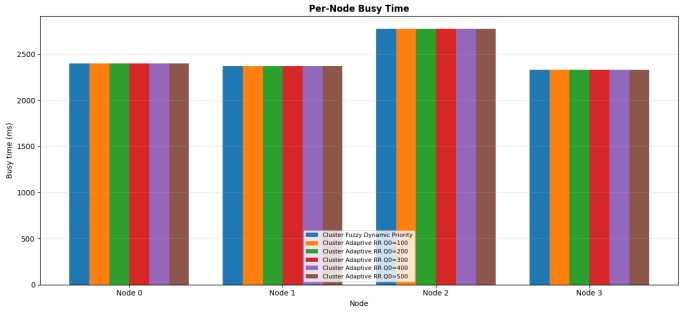



**Figure 15:** *Per-node busy time for the 4 nodes x 16 cores cluster experiment*

These results show that fuzzy dynamic priority remains effective when the model moves from a single multiprocessor queue to a node/core hierarchy. However, the advantage decreases as the system grows from 50 to 200 jobs and from 16 to 64 total cores, indicating that placement and load distribution increasingly dominate the local per-core scheduling rule.

### *4.4 Kubernetes-Like Constrained Pod Scheduling*

Table IV reports the constrained Kubernetes-like scenario with 200 pods on 4 nodes x 16 cores. Unlike the unconstrained cluster runs, node feasibility and resource waiting influence the outcome before local execution begins. Under these conditions, Adaptive RR with $Q_0 = 500$ achieves the lowest mean waiting time at 337.15 ms. Local-only fuzzy remains competitive but records 352.71 ms mean waiting time, 15.56 ms higher than the best RR setting. Two-level fuzzy improves response time slightly relative to local-only fuzzy (227.22 ms vs. 230.12 ms) and reduces total completion time (2905 ms vs. 2943 ms), but it increases mean waiting time and variance because node-level scoring can concentrate work on high-scoring feasible nodes.

| Model | Mean Wait | Std Dev | Sched Delay | Response | Total | Ctx Sw | Q Exp | Preempt |
|---|---|---|---|---|---|---|---|---|
| Local-only Fuzzy | 352.71 | 380.43 | 159.75 | 230.12 | 2943 | 170 | - | - |
| Two-level Fuzzy | 377.25 | 414.12 | 167.91 | 227.22 | 2905 | 180 | - | - |
| RR $Q_0$=100 | 372.83 | 320.00 | 183.39 | 267.08 | 2943 | 242 | 330 | 117 |
| RR $Q_0$=200 | 360.51 | 314.54 | 174.55 | 273.63 | 2912 | 212 | 256 | 87 |
| RR $Q_0$=300 | 364.58 | 322.10 | 174.71 | 280.31 | 3059 | 198 | 207 | 74 |
| RR $Q_0$=400 | 359.45 | 329.86 | 166.76 | 282.86 | 3000 | 183 | 167 | 58 |
| RR $Q_0$=500 | 337.15 | 327.15 | 157.37 | 288.22 | 3027 | 159 | 124 | 34 |

**Table 4:** *Kubernetes-like constrained pod scheduling on 4 nodes x 16 cores with 200 pods.*

Figure 16 plots the main Kubernetes-like metrics. In this constrained setting the $Q_0 = 500$ RR variant has the lowest mean waiting time and resource-wait count, while fuzzy remains stronger on response time than the RR variants

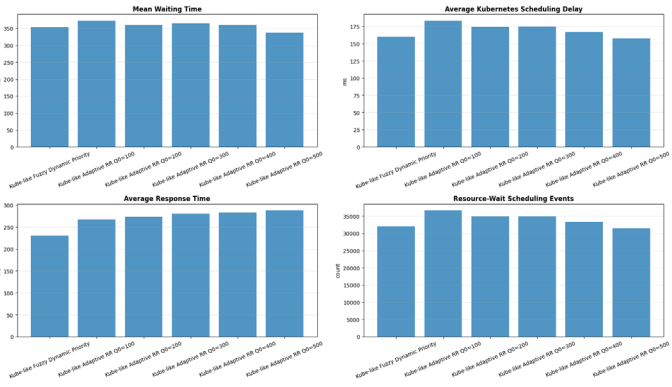



**Figure 16**. *Kubernetes-like constrained pod scheduling metrics on 4 nodes x 16 cores with 200 pods*

Table V shows the node-placement distribution for the main Kubernetes-like variants. Node 3 receives only 21 pods in all cases, reflecting the effect of feasibility constraints rather than the local scheduling policy alone. This explains why a strong per-core fuzzy rule does not automatically dominate at cluster level: a pod that waits for a feasible node or lands on a heavily loaded feasible node may already have accumulated scheduling delay before fuzzy local selection can help.

| Model | Node 0 | Node 1 | Node 2 | Node 3 | Placement interpretation |
|---|---|---|---|---|---|
| Local-only Fuzzy | 71 | 53 | 55 | 21 | Best local fuzzy wait among fuzzy variants, but uneven placement |
| Two-level Fuzzy | 51 | 61 | 67 | 21 | More scoring influence at placement level, but higher waiting variance |
| RR $Q_0$=500 | 71 | 51 | 57 | 21 | Lowest mean wait in constrained Kubernetes-like run |

**Table 5:** *Node-placement distribution in the Kubernetes-like experiment.*

The Kubernetes-like experiment therefore changes the main claim. Fuzzy Dynamic Priority is clearly better for local multicore scheduling and remains better in the unconstrained cluster-node experiments. In a constrained Kubernetes-like scheduler, however, Adaptive RR competes strongly and can outperform fuzzy on mean waiting time. A stronger fuzzy design must operate at both levels with carefully tuned node-scoring weights, not merely apply the local fuzzy rule after placement

## 5. ANALYSIS

### *5.1 Why the Fuzzy Scheduler Wins on Mean and Response*

The fuzzy scheduler's mean-waiting advantage comes

from three compounding effects. (i) The max($\mu_p$, $\mu_\beta$) rule for non-minimum processes lets either a short burst or a high priority pull a process to the front, so the scheduler captures most of SJF's short-burst gains and most of Priority's high-priority gains simultaneously. (ii) The $\mu_p + \mu_\beta$ rule for the minimum-priority process ensures that the most-starvable process is never indefinitely deferred, eliminating the long-tail waiting times that pull up Priority's mean. (iii) The global load-balancing layer keeps cores active, raising effective throughput and shrinking the queue tail.

The response-time advantage over RR is structural: under RR, the response time of any newly arrived process is at least the residual quantum of whatever is running on its core, summed with the quanta of any processes queued ahead of it. The fuzzy scheduler does not have this property a new arrival with a high $DP_i$ preempts on the next tick.

### *5.2 Why Adaptive RR Wins on Variance and Switching*

RR's strength is fairness by construction: every process receives a turn every $\lceil n/k \rceil \cdot Q\_cur$ ticks on a k-core system with n queued processes, regardless of static priority or burst time. This bounded recurrence directly compresses the waiting-time distribution, which is why all five RR settings post a lower standard deviation than fuzzy. The trade-off is that no individual process is ever preferentially advanced, so the mean waiting time cannot drop below what an oblivious rotation produces.

Larger $Q_0$ further reduces context switches because (i) each grant runs for longer before rotation, and (ii) the adaptive recomputation drifts toward higher steady-state quanta, as shown in Figure 5. The flip side is worse response time: a freshly arriving process must wait an entire quantum, which at $Q_0 = 500$ averages around 473 ms.

### *5.3 Quantum Sensitivity*

The non-monotone shape of the mean-waiting curve across $Q_0$ (480 → 500 → 470 → 471 → 448) is informative. At $Q_0 = 200$, the recomputation (4) settles near 307 ms, which is large enough to slow short-burst progress yet too small to amortize the switching cost on long bursts hence the worst mean of the sweep. At $Q_0 = 500$, the quantum stabilizes near 473 ms, large enough to push most short jobs to completion within a single grant and to amortize switches on long jobs. This finding suggests that for the studied workload mix, a larger initial quantum is preferred for adaptive RR, but the best fixed choice is workload-dependent an empirical argument in favor of always running the adaptive recomputation rather than committing to a constant.

### *5.4 Cluster-Level Interpretation*

The cluster-node experiments show that the local fuzzy advantage is real but not invariant. When node placement is unconstrained and based mainly on least-loaded assignment, fuzzy still has the best mean waiting time. As the topology grows to 64 total cores and 200 sequential jobs, the advantage over Adaptive RR $Q_0 = 500$ narrows from 51.62 ms to 26.78 ms. The reason is that scheduling quality becomes a two-part problem: node placement determines where waiting accumulates, while the local core scheduler determines how quickly queued work is served.

The Kubernetes-like experiment makes this separation explicit. Feasible-node filtering, binding, resource waiting, and node-level scoring can dominate the final metric. In that setting, local-only fuzzy improves per-node execution order but cannot undo a poor or delayed placement decision. Two-level fuzzy reduces response time and total completion time relative to local-only fuzzy, but its untuned placement score increases waiting variance. This suggests that fuzzy scheduling is promising at cluster scale only when the pod-to-node fuzzy score is weighted against load, feasibility, and resource pressure rather than treated as a direct extension of the local priority rule.

### *5.5 When to Use Which Scheduler*

The two policies occupy complementary corners of the design space:

• Fuzzy Dynamic Priority is preferable for single-host multi-core execution and lightly constrained clusters when the dominant objective is minimizing mean waiting and response time, the workload mixes short and long bursts, and there is meaningful spread in static priorities. It is also the only scheduler tested that beats SJF on waiting in the classical baseline while preventing starvation by construction.

• Adaptive Round Robin is preferable when fairness, predictable variance, and bounded context switching are more important than aggressive local priority selection. In the Kubernetes-like constrained run, the $Q_0 = 500$ configuration gives the lowest mean waiting time, showing that RR can be highly competitive when placement and resource constraints dominate local scheduling.

• Two-level Fuzzy should be treated as a tunable cluster-scheduler design rather than an automatic improvement. It is attractive when response time and makespan matter, but the node-scoring membership functions must be tuned to avoid concentrating work on a subset of feasible nodes.

### *5.6 Limitations*

The simulator abstracts memory hierarchy, cache effects, I/O blocking, NUMA topology, energy consumption, and real context-switch cost Meda and Domazet (2025). The Kubernetes-like model intentionally focuses on scheduling behavior rather than full Kubernetes operation: it does not include controllers, image-pull latency, container runtime startup, networking, probes, eviction, autoscaling, service discovery, or full pod lifecycle transitions beyond simplified pending, bound, running, and completed states. The fuzzy membership functions use fixed linear forms normalized against local maxima; richer triangular, trapezoidal, Gaussian, or workload-adaptive memberships may change the cluster-level results. Finally, the two-level fuzzy node score has not yet been optimized, so the Kubernetes-like findings should be interpreted as a first controlled simulation rather than a final scheduler design.

## 6. CONCLUSION

This paper presented a comparative study of preemptive Fuzzy Logic Dynamic Priority and Adaptive Round Robin scheduling across three increasingly realistic settings: a 16-core local multiprocessor, an unconstrained cluster-node model, and a Kubernetes-like constrained pod-scheduling model. In the local 50-process experiment, fuzzy dynamic priority is the strongest policy for average waiting and response time, reducing mean waiting to 373.98 ms compared with 447.74 ms for the best Adaptive RR setting and improving response time by 160.94 ms.

In the cluster-node extensions, fuzzy remains better on mean waiting time for both the 4-node x 4-core/50-process and 4-node x 16-core/200-process scenarios, but its advantage narrows as placement effects become stronger. In the Kubernetes-like constrained scenario, Adaptive RR $Q_0 = 500$ achieves the best mean waiting time, while fuzzy remains competitive on response time and two-level fuzzy improves response and makespan relative to local-only fuzzy. Therefore, the final conclusion is not that fuzzy universally dominates RR, but that fuzzy dynamic priority is highly effective for local and lightly constrained multi-core scheduling, whereas Kubernetes-like clusters require scheduler-level fuzzy scoring to be carefully tuned alongside node feasibility and resource pressure.

Future work should optimize the two-level fuzzy node score, test nonlinear membership functions, include additional Kubernetes lifecycle elements, evaluate heterogeneous node capacities, and validate the scheduler on a real container-orchestration testbed. A promising direction is a hybrid policy that uses fuzzy logic for pod-to-node scoring and local dynamic priority, while retaining RR-style bounded service guarantees for fairness-sensitive queues.

## 7. REFERENCES

A. Silberschatz, P. B. Galvin, and G. Gagne, Operating System Concepts, 10th ed. Wiley, 2018.

W. Stallings, Operating Systems: Internals and Design Principles, 9th ed. Pearson, 2017.

L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, no. 3, pp. 338–353, 1965.

X. Li, Y. Zhang, and J. Wang, "Fuzzy logic-based dynamic priority scheduling for real-time systems," IEEE Trans. Fuzzy Systems, vol. 28, no. 4, pp. 712–725, 2020.

R. Kumar and S. Singh, "Multi-core load balancing using fuzzy inference system," Int. J. Parallel Programming, vol. 49, pp. 234–251, 2021.

A. S. Tanenbaum and H. Bos, Modern Operating Systems, 5th ed. Pearson, 2022.

R. J. Matarneh, "Self-adjustment time quantum in round robin algorithm depending on burst time of the now running processes," American J. Applied Sci., vol. 6, no. 10, pp. 1831–1837, 2009.

S. M. Mostafa, S. Z. Rida, and S. H. Hamad, "Finding time quantum of round robin CPU scheduling algorithm in general computing systems using integer programming," Int. J. Research and Reviews in Applied Sci., vol. 5, no. 1, pp. 64–71, 2010.

P. S. Varma, "A finest time quantum for improving Shortest Remaining Burst Round Robin (SRBRR) algorithm," J. Global Research in Computer Sci., vol. 4, no. 3, 2013.

A. Noon, A. Kalakech, and S. Kadry, "A new round robin based scheduling algorithm for operating systems: dynamic quantum using the mean average," Int. J. Computer Sci. Issues, vol. 8, no. 3, pp. 224–229, 2011.

Kubernetes Documentation, "Kubernetes Scheduler," https://kubernetes.io/docs/concepts/scheduling-eviction/kube-scheduler/. Accessed: May 14, 2026.

Kubernetes Documentation, "Pod Lifecycle," https://kubernetes.io/docs/concepts/workloads/pods/pod-lifecycle/. Accessed: May 14, 2026.

Kubernetes Documentation, "Assigning Pods to Nodes," https://kubernetes.io/docs/concepts/scheduling-eviction/assign-pod-node/. Accessed: May 14, 2026.

Kubernetes Documentation, "Taints and Tolerations," https://kubernetes.io/docs/concepts/scheduling-eviction/taint-and-toleration/. Accessed: May 14, 2026.

S. Meda and E. Domazet, "Computer architecture optimization techniques for AI workloads," in Bridging Horizons in Artificial Intelligence, Robotics, Cybersecurity, Smart Cities, and Digital Economy (ICITTBT 2024), K. Dhoska and E. Spaho, Eds. Cham, Switzerland: Springer Nature, 2025, ch. 25. doi: 10.1007/978-3-031-72029-1_25.

# PUBLIC FINANCIAL GOVERNANCE IN ALBANIA — OVERVIEW

***Endri Balla[1], Albana Demi (Mosho) [2]*, Linert Lireza[3]***

*1 "Aleksandër Moisiu" University of Durrës, Albania*
*2 "Aleksandër Moisiu" University of Durrës, Albania*
*3 "Aleksandër Moisiu" University of Durrës, Albania*

## Abstract

This study assesses how accounting transparency and auditing influence the reliability of financial performance in Albania's public financial system. It explores the relationship between public sector accounting practices, audit effectiveness, and financial governance. The authors highlight the importance of transparency and accountability, concluding that transparent accounting directly enhances the reliability of financial reports. They recommend modernizing practices to foster sustainable governance and economic growth. The research tests three hypotheses: (1) accounting transparency improves financial report reliability, (2) effective auditing reduces irregularities and damage, and (3) implementing audit recommendations enhances financial performance and accountability. Methodologically, the study employs a quantitative, descriptive approach, analyzing public institution publications, legal frameworks, audit reports, and data from a structured questionnaire completed by 420 public sector professionals in finance, accounting, auditing, and administration. Data analysis involved descriptive statistics, comparative methods, and document review to identify trends and correlations among transparency, audits, and financial indicators. The findings reveal a significant positive impact of transparency and auditing on financial reliability. Most respondents view transparent accounting as vital for accountability and reducing irregularities. The analysis also shows that higher implementation of audit recommendations correlates with better financial management and less economic damage to the state. The study emphasizes that enhancing transparency, audit effectiveness, and the application of audit recommendations are crucial for strengthening public financial governance and accountability in Albania. It contributes empirical evidence and policy suggestions to improve financial management in the public sector.

JEL Classification: H1, H83, G1, G2, G18, M42

**Keywords:** Auditing, Internal Control, Accounting, Public Finance, Legislation, Financial Statement, Eu Integration, Transparency.

## 1. INTRODUCTION

The public financial system in Albania has undergone significant transformations during the last decade, presenting both challenges and opportunities related to technological innovation, regulatory reforms, and the increasing demand for financial transparency and accountability. Public accounting and auditing represent the two fundamental pillars of financial reporting within the public sector, playing a crucial role in strengthening fiscal discipline, improving decision-making processes, and ensuring the credibility and reliability of public financial information. Effective public accounting systems contribute to the accurate recording, analysis, and reporting of financial transactions, while public auditing functions as an independent mechanism for evaluating financial operations, mitigating the risks of mismanagement, fraud, and misuse of public resources.

Despite continuous reforms in the field of public financial management, Albania continues to face challenges associated with fragmented reporting systems, limited institutional capacities, and the approximation of accounting and auditing standards with European Union requirements. These challenges underline the importance of strengthening public financial governance through improved accounting transparency, effective audit mechanisms, and enhanced internal control systems. In this context, the interaction between public accounting and auditing becomes essential for promoting transparency, accountability, sustainability, and institutional efficiency within the Albanian public sector.

The legal and institutional framework governing public sector financial accounting and reporting in

*Corresponding author: Albana Demi (Mosho) , albanamosho@uamd.edu.al

Albania remains relatively complex and fragmented. The Albanian Ministry of Finance plays a central role in public financial management through fiscal policy formulation, budget preparation, debt management, internal financial control, and the approval of financial management methodologies, reporting standards, and accounting procedures for public sector institutions. In parallel, the role of the Supreme Audit Institution of Albania, Supreme Audit Institution of Albania (ALSAI/KLSH), has become increasingly important in enhancing transparency, accountability, and communication with stakeholders through the publication of audit reports, statistical analyses, recommendations, and press releases regarding public financial performance. The study is developed within the broader context of Albania's economic development and European integration process. Albania has achieved upper-middle-income status and continues to advance toward European Union membership following the formal launch of accession negotiations in July 2022. Economic growth during the period 2024–2025 has been supported by domestic demand, tourism, and construction activities, while international institutions have emphasized the necessity of further reforms in governance, public financial management, anti-corruption policies, and institutional modernization. The implementation of the National Reform Agenda 2024–2027 and Albania's participation in the European Union Reform and Growth Facility further highlight the strategic importance of strengthening transparent and accountable financial systems. Public accounting serves the public interest through the provision of accounting, auditing, taxation, and advisory services that enhance the reliability and integrity of financial information. In a complex and increasingly digital global economy, public accountants, auditors, policymakers, investors, and law enforcement institutions rely on accurate financial reporting to support financial stability and economic governance. Consequently, public accounting is strictly regulated and based on the principles of independence, integrity, objectivity, professional competence, and ethical conduct. At the same time, technological developments such as artificial intelligence, blockchain technologies, and digital financial systems are transforming accounting and auditing practices, requiring professionals to develop new technical competencies and adaptive institutional capacities.

This research paper examines the interaction between public accounting and auditing systems in Albania from both theoretical and practical perspectives, focusing on their impact on transparency, accountability, financial performance credibility, and sustainable public financial governance. The study covers the period 2013–2023 and analyzes the relationship between accounting transparency, audit effectiveness, internal control quality, and public financial performance. Particular attention is given to the role of accounting and auditing in fraud detection and prevention, institutional efficiency, and the implementation of public policies and programs. Transparent financial systems are considered essential for increasing investor confidence, supporting economic growth, and improving Albania's convergence with European Union standards. The study is based on a quantitative and descriptive research methodology. Data were collected over a two-month period through an anonymous survey distributed online via Google Forms, email communication, and in-person administration. The research applies purposive sampling and includes respondents employed in finance, accounting, auditing, and public administration functions within Albanian public institutions. The analytical framework combines descriptive statistical analysis with institutional document analysis and is supported by auditing theory, stakeholder theory, and institutional theory. These theoretical approaches help explain the role of independent auditing in reducing information asymmetry, the broader responsibility of public accountants toward society, and the influence of institutional regulations and professional standards on accounting practices.

The findings of the study indicate that accounting transparency and auditing positively affect the credibility and reliability of financial performance and are essential for the long-term development of the Albanian economy. The study concludes that strengthening transparent financial systems, improving international cooperation, and enhancing the professional capacities of accountants and auditors are critical for achieving sustainable economic development, effective public financial governance, and successful European integration.

## 2. LITERATURE REVIEW

Public accounting is essential to the economy and faces not only challenges but also opportunities due to globalization, technological advances, and regulatory changes. The academic and institutional literature on public financial governance in Albania spans theoretical contributions, empirical studies, and policy evaluations that collectively underscore the need for stronger coordination between accounting practices and audit mechanisms.

In the 1990s, Albania faced significant challenges due to public reforms, which inhibited economic growth and led to a civil war in 1997. This instability highlighted the institutions' lack of preparedness and reflected deeper cultural and psychological issues stemming from political turmoil. Despite receiving international financial aid, Albania continues to struggle with public sector financial management, as noted in recent European Commission reports, indicating persistent anomalies in fund allocation influenced by cultural factors and the resistance of

political actors (Blerina Sadiku & Majlinda Velcani, 2021).

The theory and practice of public finance focus on fulfilling social needs, safeguarding public rights, and ensuring the production of quality public goods in adequate amounts within the economic domain. The distinction between public and private sector accounting is particularly relevant in the Albanian context. Albania is at risk for money laundering due to corruption, organized crime, and legislative gaps, resulting in its placement on the FATF gray list for increased monitoring as of March 2022 (Demi (Mosho) A., 2022). This risk has intensified in the context of the real estate boom in Tirana, where construction sector transactions have been used to launder illicit proceeds. In 2024, Albanian authorities seized approximately €50 million in assets linked to money laundering in the capital, exposing the scale of the problem and reinforcing the need for robust financial auditing. The key differences in accounting for financial statements between public and private sector entities stem from the public sector's focus on fund management and achieving value for money rather than profit generation. While the foundational accounting principles remain consistent, the presentation of statements and accounting treatments differ significantly. Accountants face procedural choices in financial reporting between various alternatives (Mosho, 2016). Input from auditors, accountants, and financial officers is crucial for policymakers to assess the effectiveness of practices and economic impact, highlighting the importance of responsibility, fairness, and ethical standards in public asset management for sustainable development and prioritization of public services (Demi A., 2025). The SIGMA Monitoring Report on Public Administration in Albania (OECD/SIGMA, 2024) confirms that while internal audit functions are well-established and the independence of the State Supreme Audit Institution (SSAI) is ensured, significant challenges persist, including a high level of tax arrears (42%), inconsistent compliance with commitment controls, and a low implementation rate for both internal and external audit recommendations. EU integration in Albania necessitates major reforms. While the economy shows potential with decreasing poverty levels, challenges like demographic shifts, climate risks, and structural issues persist. The Medium-Term Revenue Strategy (2024–2027) aims to boost revenue collection through initiatives such as formalizing short-term rentals and implementing a carbon tax, which could raise revenues by up to 0.9% of GDP (World Bank, 2025). Budget priorities remain constrained, with over one-third allocated to pensions and transfers.

Albania has achieved significant progress in public financial management reforms over recent decades. The Public Financial Management Strategy 2023-2030, adopted by Council of Ministers Decision No. 390 on 12 June 2024, provides a strategic framework for PFM reform, specifying objectives and measurable performance indicators. However, the OECD/SIGMA assessment (2024) notes that more than half (52%) of outcome-level indicators lack baseline values, and several targets are described only as directional trends, which undermines effective monitoring.

Albania is implementing structural reforms aimed at promoting equitable economic growth, improving governance, and enhancing public services. The government focuses on fostering faster growth through regional integration and market diversification, with priority areas including territorial devolution, financial stability, energy, social assistance, and fiscal stability. Albania has also adopted the Anti-Corruption Strategy 2024-2030 and the Public Administration Reform Strategy 2025–2030, offering a coherent roadmap for institutional reform (European Commission, 2025). Recent amendments to the Public Procurement Law (Law No. 16/2024) introduced electronic communication at all stages, while the National Strategy for Public Procurement 2024-2030 aims to enhance integrity (Demi Mosho A., 2023; Transparency International, 2025). In July 2024, the Albanian National Assembly approved the establishment of a Sub-Committee for Public Sector Audit dedicated to the work of the Albanian Supreme Audit Institution (ALSAI). The sub-committee's mandate includes the examination of audit reports, monitoring the implementation of recommendations (follow-up), and reporting findings to the Assembly. ALSAI has also established a Directorate for Parliamentary Relations to manage communication with this sub-committee and other relevant parliamentary committees, strengthening the role of external audit in democratic oversight (Transparency International, 2025).

According to Prof. Linert Lireza, environmental issues in the Balkans affect security through economic insecurity, human displacement, health concerns and energy security. Regional cooperation and international assistance are essential to address these problems, with a focus on sustainable development (Lireza, 2023). These environmental pressures also impose fiscal costs on public institutions that must be accounted for in long-term financial planning.

Financial depth negatively impacts wealth distribution, despite its potential to increase investment. To improve income redistribution, it is essential to combine financial depth with access to financial technology (fintech) and supportive credit policies for the poor. GDP growth positively impacts consumption, employment, and wage competitiveness, which can reduce income inequality. Nations with lower inequality tend to exhibit higher levels of education, stable inflation, and increasing government spending (Imeraj et al., 2025). Public expenditures in Albania are positively correlated with the nation's economic growth (Demi et al., 2021). Grossi et al. (2023)

argue that the boundaries of public sector auditing are changing globally, as new mandates extend beyond financial compliance to include performance, environmental, and digital governance audits. These global trends are gradually manifesting in the Albanian context, particularly as ALSAI expands its audit coverage to include cybersecurity and data protection in the public sector. Hoxhaj (2018) examined Albania's public sector accounting reform and concluded that the transition from cash-based to full accrual accounting remains the core challenge, as it requires significant capacity building, system upgrades, and legal amendments. While the Albanian government has adopted International Public Sector Accounting Standards (IPSAS) as a long-term target, the pace of implementation has been constrained by institutional capacity. Bello (2013) similarly highlighted the critical role of audit committees in enhancing accountability in Albanian public sector institutions, noting that their effectiveness depends on the independence and professional competence of members.

## 3. METHODOLOGY AND DATA

### *3.1 Research Design and Approach*

This study employs a mixed-methods research design, combining quantitative survey data with qualitative analysis of institutional documents, legal frameworks, and secondary statistical sources. The research aims to identify the main difficulties and challenges in the practical implementation of changes in the field of Public Accounting and Auditing, and to examine the role and effectiveness of employees in managing the process. Based on the data collected through questionnaires, the study provides concrete recommendations for improving the legal and administrative framework, with the aim of building a more transparent, effective and fair public financial governance system in Albania.

### *3.2 Data Sources*

For this research, primary data and official statistics were collected from the websites of public institutions, including: INSTAT (the Institute of Statistics of Albania), the Ministry of Finance and Economy, and the Albanian Supreme Audit Institution (KLSH/ALSAI). INSTAT offers transparent, neutral, and timely statistics that assist users in evaluating the progress of transformation processes. Secondary data was drawn from reports by the World Bank, the International Monetary Fund (IMF), the European Commission, and the OECD/SIGMA. The study covers the period 2013–2023, which encompasses a decade of institutional reforms in the context of Albania's EU accession aspirations. The direct method was used to detect economic agents in the informal economy, though this approach can be costly and risks inaccurate reporting due to lack of confidence and issue-sensitive participation. Tax audits were also considered as an estimation tool by comparing income declared for tax purposes with those measured by selection controls.

### *3.3 Survey Instrument*

The primary data collection instrument was a structured questionnaire comprising 25 questions distributed to employees in public institutions. A total of 500 copies were distributed, of which 420 were returned, yielding a response rate of 84 percent. The sample consisted of 88.9% women and 11.1% men. The questionnaire targeted four professional groups: public finance officers, auditors, accountants, and other public sector employees. It was administered through an anonymous online platform (Google Forms), by email, and through in-person distribution

### *3.4 Research Hypotheses*

The role and interaction of public accounting and auditing as the main pillars of public financial governance in Albania are examined through the following hypotheses and research questions. Together, the hypotheses and questions provide a structured analytical framework to assess existing challenges, evaluate the impacts of reform, and generate evidence-based recommendations for strengthening public financial governance in Albania

H1: Improving the transparency and reliability of financial information has a positive impact on strengthening public accounting standards and public financial governance in Albania.

H2: Effective public sector auditing (internal and external) significantly contributes to increasing accountability and control over public finances in Albania.

H3: In public financial governance in Albania, the interaction between public accounting systems and independent audit mechanisms reduces financial mismanagement and increases public trust.

### *3.5 Research Questions*

Q1: How do public accounting practices impact the transparency and decision-making of public finance governance in Albania?

Q2: What is the role of public auditing in guaranteeing compliance and accountability in Albania's public fund management?

Q3: To what extent does coordination between public accounting and auditing systems affect the effectiveness of public financial governance in

Albania?
Even after the ongoing implementation of public financial management reforms, there are still limited empirical studies on the interaction between public accounting and auditing systems and their impact on transparency, accountability, and institutional performance in Albania. This scholarly study aims to address this gap by providing evidence-based analysis and recommendations aimed at strengthening public financial governance and institutional accountability.

## 4. RESEARCH RESULTS AND DISCUSSION

### *4.1 Macroeconomic and Fiscal Context*

Public finances contribute to economic growth but pose burdens on future generations due to repayment responsibilities. In Albania, public finance management has improved substantially through initiatives such as an electronic treasury system, program-based budgeting frameworks, and reforms in public debt management and monitoring. The IMF (2025) notes that while progress has been made in strengthening governance and reducing fiscal risks, the composition and quality of public spending continues to require improvement, particularly regarding state-owned enterprises (SOEs) and the management of fiscal risks. The Albanian economy grew at 3.9 percent in 2023 and approximately 4.0 percent in the first half of 2024, supported by services, tourism, and construction activity. However, the IMF has cautioned that sustaining this growth requires addressing structural weaknesses in public financial management, reducing informality — which remains a key factor limiting tax revenue collection — and advancing governance reforms linked to EU accession. Revenue collection remains below the EU average due to high informality, broad exemptions, and enforcement gaps.
Medium-term fiscal projections indicate that Albania's public debt ratio is expected to decline to around 52 percent of GDP by 2029. The SIGMA monitoring report (OECD/SIGMA, 2024) highlights that while the internal audit framework aligns with international standards, the weak credibility of medium-term fiscal forecasts and a high level of tax arrears — standing at approximately 42 percent — represent persistent governance weaknesses. Compliance with commitment controls remains inconsistent, leading to payment arrears in many government bodies.

### *4.2 Audit Activity Trends: Evidence from ALSAI (2013–2023)*

The audit activity data published on the ALSAI website (www.klsh.gov.al) provides a comprehensive overview of audit trends from 2013 to 2023. The data in Table 1 and Graphic 1 illustrates the evolution of audit coverage, staffing, and financial impact across a decade of reform.
The analysis reveals that the audit process has become more robust in recent years, with more auditors engaged, a higher volume of recommendations, and increasingly significant economic damage identified.

The total number of audits varies each year and has seen a significant increase in recent years. In 2014 it reached a peak of 192 audits. A noticeable decrease occurred in 2020 — down to 73 audits — as a direct result of the COVID-19 global pandemic, which severely disrupted audit activity across all public institutions.

In 2023, 160 audits were performed, reflecting a fairly stable level of activity in recent years. The number of audited institutions also increased overall, reaching a peak of 240 in 2016.

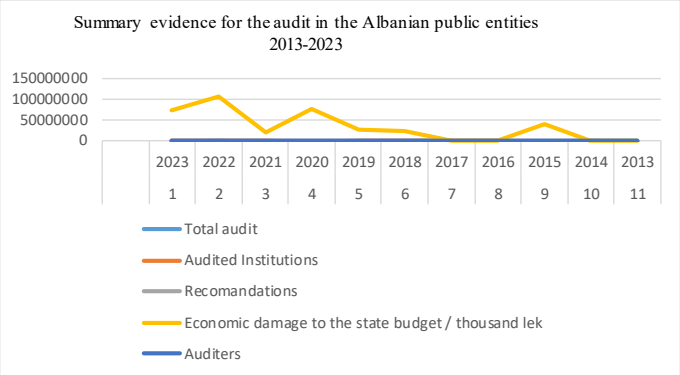
The total number of recommendations made has fluctuated considerably over the decade, reaching 5,759 in 2023 — a significant increase after 2021 that results from more comprehensive auditing in prior years. Even though there were fewer audits performed in 2022 compared to earlier periods, the amount of economic damage identified was among the highest on record. This implies that year's audits had a particularly significant impact, possibly reflecting improved auditor training and the detection of more severe compliance failures.

There is a clear positive correlation between the number of auditors and the volume of audits and recommendations produced. Significant discoveries in terms of economic impact and recommendations have been observed in the most recent years (2022 and 2023), which may indicate increased audit efforts or the discovery of more serious problems inside audited entities. These findings align with the European Commission's 2025 Albania Progress Report, which notes that while audit findings are formally accepted, the implementation rate of recommendations remains low, suggesting that the follow-up mechanism requires strengthening.

In 2024, the Albanian Parliament's newly established Sub-Committee for Public Sector Audit — created in July 2024 — strengthened parliamentary oversight of ALSAI's work. ALSAI has conducted joint audits with counterpart institutions from the Netherlands (Algemene Rekenkamer) and other European SAIs within the EUROSAI framework, improving audit quality through the application of international methodologies. In 2023, ALSAI extended its audit scope to include cybersecurity and data protection among 152 audited public institutions, including the Bank of Albania, the Customs Directorate, and the Tax Directorate.

| No. | Year | Total audit | Audited Institutions | Recomandations | Economic damage to the state budget / thousand lek | Auditers |
|---|---|---|---|---|---|---|
| 1 | 2023 | 160 | 195 | 5759 | 72,386,606 | 165 |
| 2 | 2022 | 177 | 205 | 5420 | 107,488,177 | 163 |
| 3 | 2021 | 158 | 205 | 4251 | 19,158,985 | 130 |
| 4 | 2020 | 73 | 91 | 1430 | 74,490,878 | 134 |
| 5 | 2019 | 160 | 192 | 3441 | 27,304,777 | 148 |
| 6 | 2018 | 171 | 206 | 4232 | 24,093,742 | 128 |
| 7 | 2017 | 156 | 207 | 1348 | 11.500.000 | 118 |
| 8 | 2016 | 154 | 240 | 1515 | 25.484.000 | 119 |
| 9 | 2015 | 158 | 110 | 1590 | 37,743,113 | 116 |
| 10 | 2014 | 192 | 130 | 924 | 11 085 084 | 111 |
| 11 | 2013 | 191 | 110 | 805 | 15 077 225 | 106 |
| **TOTALI** | | **1750** | **1891** | **30715** | **362666278** | **1438** |

**Table 1:** *Summary evidence for audit in Albanian public entities, 2013–2023*




**Graphic 1:** *Summary evidence for the audit in the Albanian public entities 2013-2023*

### *4.3 Public Accounting Standards and Practices*

Public Accounting systematically records and reports financial transactions and is essential in public service, ensuring the proper use of citizens' funds and supporting policy goals. Public institutions, through comprehensive accounting practices, maintain fairness in asset management, build credibility and minimize the risk of asset misuse. The Public Accounting system ensures the protection of public assets, proper authorization and recording of transactions, and is crucial for managing the financial and operational processes of public institutions. It relies on administrative procedures and financial regulations to facilitate risk management, allowing institutions to identify and reduce risks effectively.
Public accounting extends to advisory and consulting services, including financial restructuring, and encompasses auditing and assurance to verify compliance of financial statements with SKK (National Accounting Standards) and SNK (International Accounting Standards) along with ethical tax planning and compliance services. The Albanian Ministry of Finance has approved guidelines for improving the monitoring and reporting of internal audit recommendations, including a standardized methodology for financial statements, audit trails, and accounting procedures. The legislation on internal audit practice formally aligns with international standards; however, audit opinions on the effectiveness of internal control systems across budget entities are not consistently issued. The Public Financial Management Sectoral Strategy 2023-2030 of Albania aims to enhance accountability and transparency through better financial and non-financial performance reporting in line with international standards. It covers key areas including external audit, internal audit and financial control, public procurement, and anti-fraud coordination. The strategy envisages a shift from cash-based accounting towards accrual-based accounting, a transition that is consistent with IPSAS standards and EU Directive 2011/85/EU on requirements for national budgetary frameworks. However, as Hoxhaj (2018) noted, moving to full accrual accounting remains a long-term challenge requiring sustained investment in human and technical resources.

### *4.4 Questionnaire Results: Practitioners' Perspectives*

The authors prepared a questionnaire with the aim of providing information on the importance and valuable insights into public accounting and auditing practices in Albania. The questionnaire has in total 10 questions to providing and focusing on public accounting and auditing practices, it has been distributed a total of 500 copies to public institution employees we only collected 420.

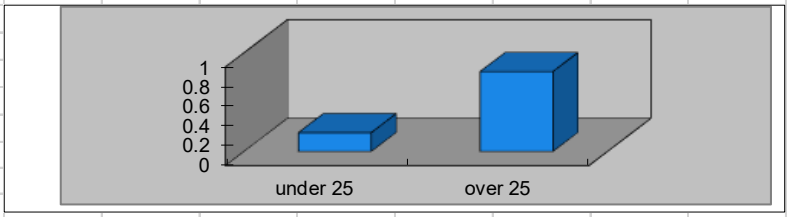
The questionnaire examines the changes in audit coverage, economic damage to the state budget, the effectiveness of auditors, institutional compliance, and long-term recommendations for improving

public financial management. It also explores the relationship between auditor numbers and audits conducted annually, the challenges faced by public institutions in meeting audit recommendations, and the effectiveness of audits in promoting transparency and accountability.

The questionnaire also identifies the most effective recommendations and the mechanisms for tracking implementation. The overall auditing framework has evolved over time to address emerging.

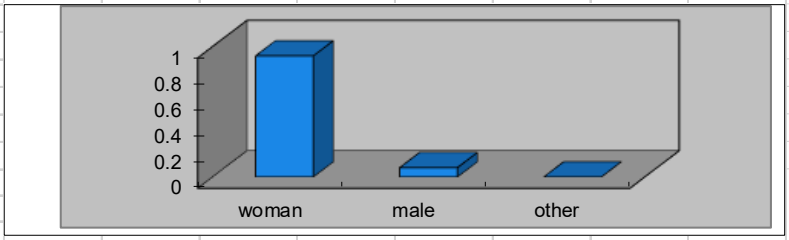
Question 1. Age

| Answer | Percent(%) | Count(N) |
|---|---|---|
| under 25 | 19.05% | 66 |
| over 25 | 80.95% | 354 |
| Total | 100.00% | 420 |
| Mean | 1.81 | |
| Confidence Interval @ 95% | [1.689 - 1.930] | |
| Standard Deviation | 0.397 | |
| Standard Error | 0.061 | |



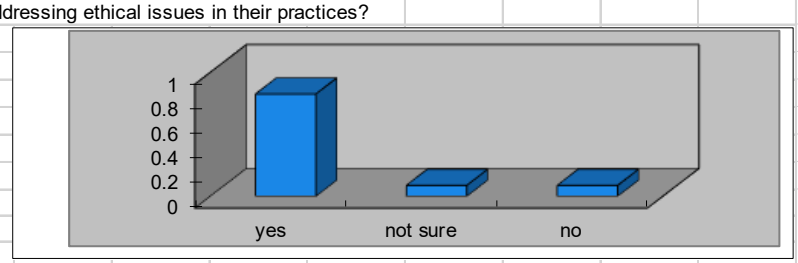

Question 2. Gender

| Answer | Percent(%) | Count(N) |
|---|---|---|
| woman | 92.86% | 390 |
| male | 7.14% | 30 |
| | | 0 |
| Total | 100.00% | 420 |
| Mean | 1.07 | |
| Confidence Interval @ 95% | [0.993 - 1.150] | |
| Standard Deviation | 0.261 | |
| Standard Error | 0.040 | |



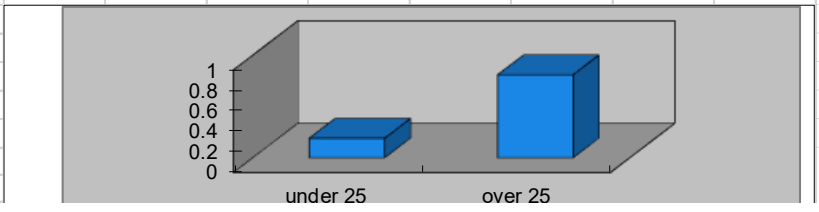

Q3. Do you believe that public accounting reports are adequately addressing ethical issues in their practices?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| yes | 82.50% | 330 |
| not sure | 8.75% | 35 |
| no | 8.75% | 35 |
| Total | 100.00% | 420 |
| Mean | 1.18 | |
| Confidence Interval @ 9 | [1.056 - 1.294] | |
| Standard Deviation | 0.385 | |
| Standard Error | 0.061 | |



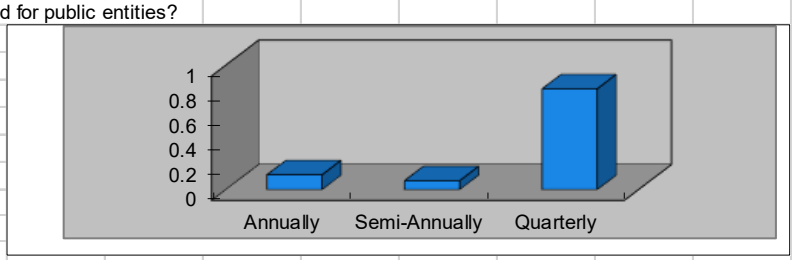

Q1. Age

| Answer | Percent(%) | Count(N) |
|---|---|---|
| under 25 | 19.05% | 66 |
| over 25 | 80.95% | 354 |
| Total | 100.00% | 420 |
| Mean | 1.81 | |
| Confidence Interval @ 95% | [1.689 - 1.930] | |
| Standard Deviation | 0.397 | |



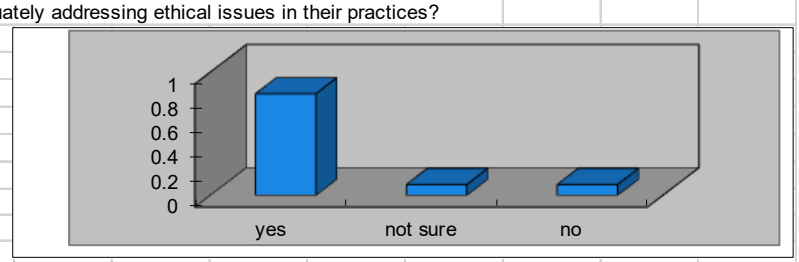

Question 5.How frequently do you believe audits should be conducted for public entities?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| Annually | 11.90% | 50 |
| Semi-Annually | 7.14% | 30 |
| Quarterly | 80.95% | 340 |
| Total | 100.00% | 420 |
| Mean | 2.69 | |
| Confidence Interval @ 9 | [2.485 - 2.896] | |
| Standard Deviation | 0.680 | |
| Standard Error | 0.105 | |



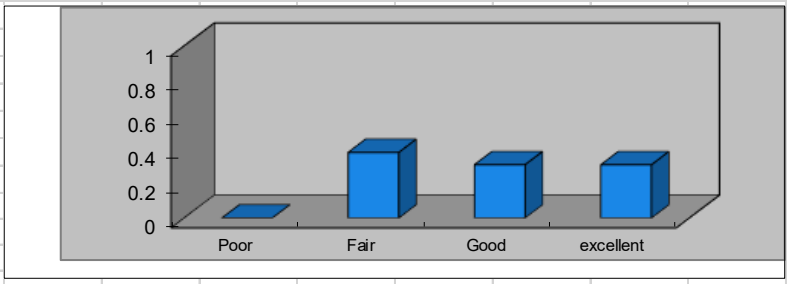
Question 6. Do you believe that public accounting reports are adequately addressing ethical issues in their practices?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| yes | 82.50% | 330 |
| not sure | 8.75% | 35 |
| no | 8.75% | 35 |
| Total | 100.00% | 420 |
| Mean | 1.18 | |
| Confidence Interval @ 9 | [1.056 - 1.294] | |
| Standard Deviation | 0.385 | |
| Standard Error | 0.061 | |



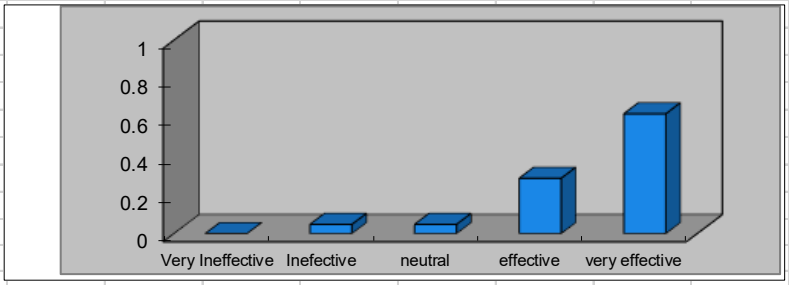

Question 7. How would you rate the overall quality of services provided by public accounting profesionals?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| Poor | 0.00% | 0 |
| Fair | 38.10% | 160 |
| Good | 30.95% | 130 |
| excellent | 30.95% | 130 |
| Total | 100.00% | 420 |
| Mean | 2.93 | |
| Confidence Interval @ 9 | [2.675 - 3.182] | |
| Standard Deviation | 0.838 | |
| Standard Error | 0.129 | |



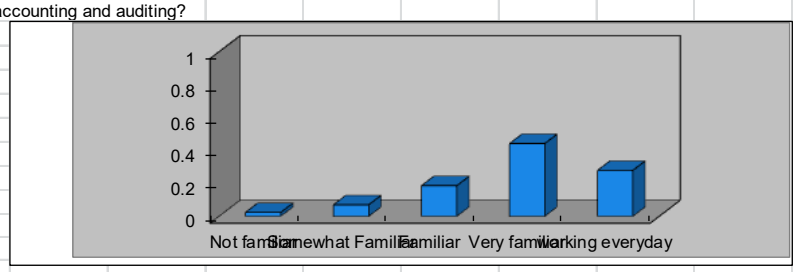

Question 8. How effective do you find current auditing standards and practices in ensuring financial transparency?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| Very Ineffective | 0.00% | 0 |
| Inefective | 4.76% | 20 |
| neutral | 4.76% | 20 |
| effective | 28.57% | 120 |
| very effective | 61.90% | 260 |
| Total | 100.00% | 420 |
| Mean | 4.48 | |
| Confidence Interval @ 95% | [4.233 - 4.719] | |
| Standard Deviation | 0.804 | |
| Standard Error | 0.124 | |



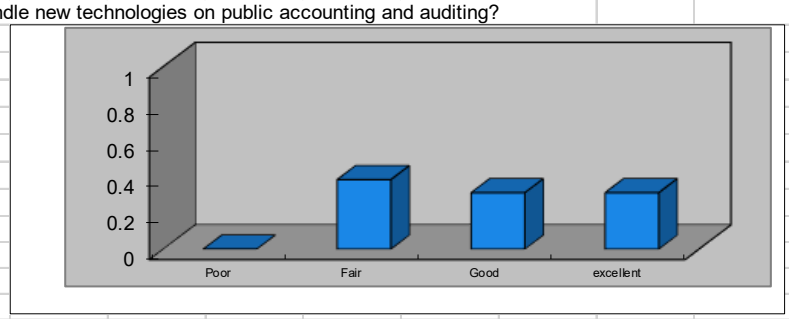

Question 9. How familiar are you with the latest regulations impacting public accounting and auditing?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| Not familiar | 2.33% | 10 |
| Somewhat Familiar | 6.98% | 30 |
| neutral | 18.60% | 80 |
| Very familiar | 44.19% | 190 |
| working everyday | 27.91% | 110 |
| Total | 100.00% | 420 |
| Mean | 3.88 | |
| Confidence Interval @ 95% | [3.591 - 4.177] | |
| Standard Deviation | 0.981 | |
| Standard Error | 0.150 | |




Question 10. Do you think the public institutions are prepared to handle new technologies on public accounting and auditing?

| Answer | Percent(%) | Count(N) |
|---|---|---|
| Not prepared | 0.00% | 0 |
| Somewhat prepared | 38.10% | 160 |
| Well prepared | 30.95% | 130 |
| Very well prepared | 30.95% | 130 |
| Total | 100.00% | 420 |
| Mean | 2.93 | |
| Confidence Interval @ 9 | [2.675 - 3.182] | |
| Standard Deviation | 0.838 | |
| Standard Error | 0.129 | |



The survey, distributed and completed by 420 respondents, highlights key insights into public accounting and auditing practices in Albania.
The majority of respondents (80.95%) are over 25 years old, predominantly women (92.86%), indicating a professionally mature and gender-unbalanced workforce. Almost half (47.62%) have 4 to 10 years of experience and are divided into auditors (38.10%) and accountants (30.95%), who represent the largest professional roles. A significant percentage of 80.95% favor quarterly audits for public institutions, reflecting a desire for increased oversight. Regarding ethical issues, 82.50% believe that current reporting practices adequately address concerns. Responses on service quality are mixed, with 38.10% rating it as fair. The majority (61.90%) see audit standards as very effective in ensuring transparency. Awareness of recent regulations is moderate, with 44.19% feeling very familiar. Preparedness for technological trends varies, with 38.10% feeling somewhat prepared. Overall, there is a need to improve service quality and adapt to technology.

The questionnaire results confirm a broadly positive but cautious assessment of current public accounting and auditing practices in Albania. A majority of respondents rated the overall quality of services provided by public accounting professionals as "good" or "excellent." However, a significant share rated the effectiveness of current accounting standards and practices in ensuring financial transparency as "neutral" or "ineffective," pointing to a persistent gap between formal standards and practical implementation. The reliability of accounting information for managing public revenues and expenditures was rated as "moderately reliable" or "highly reliable" by the majority, suggesting progress in data quality.
Regarding audit effectiveness, most respondents found public auditing "moderately effective" or "effective" in identifying weaknesses in public financial management. The extent to which audit findings influence financial management decisions was rated "moderately" to "significantly" by most respondents. Audit recommendations were reported as being "sometimes" or "often" followed and implemented — a finding that aligns with the SIGMA (2024) assessment that audit recommendation implementation rates remain low across the Albanian public sector. On coordination between accounting and auditing units, most respondents rated it as "moderate" or "weak," indicating significant room for improvement in inter-institutional cooperation.

With respect to the impact of accounting–auditing interaction on public financial governance, a majority of respondents agreed or strongly agreed that such interaction enhances transparency in public finance and improves accountability in the use of public funds. The open-ended responses identified as the main challenges: (1) fragmented institutional coordination; (2) inadequate professional training; (3) limited technological infrastructure for financial reporting; and (4) political interference in the audit process. The most frequently recommended improvements included: the development of integrated financial management information systems; standardized reporting formats aligned with EU requirements; increased independence of internal audit units; and greater parliamentary engagement in monitoring audit findings. Improving the transparency and reliability of financial information positively impacts public accounting standards and governance in Albania, with partial support for:

H1. Evidence shows high agreement on ethical reporting (82.50%) and effective audit standards (61.90%), but discrepancies in service quality persist.

H2 receives strong support, highlighting that effective public sector auditing enhances accountability and financial control, demonstrated by a strong preference for quarterly audits (80.95%).

H3, has partial support regarding the interaction between public accounting and independent audits, as while ethical reporting is perceived positively, challenges remain in efficiency and digital readiness.

This research study highlights the importance of transparency and effective auditing in improving public financial governance outcomes.
The analysis relies on percentages and descriptive summaries rather than regression, correlation, chi-square, or other hypothesis testing methods. The wording throughout the document is softened by causal expressions such as “significantly contributes” or “positively impacts” to descriptive language such as “respondents perceive” or “findings suggest." H1: This study explores whether respondents perceive public sector auditing as contributing to financial accountability and control. and the results indicate that the survey responses suggest that respondents generally perceive auditing as effective for the public sector.

## 5.CONCLUSION AND POLICY RECOMMENDATIONS

This research paper has comprehensively examined the role and interaction of public accounting and auditing as the main pillars of public financial governance in Albania. The analysis of accounting data and audit statistics from 2013 to 2023 confirms that these two mechanisms are critically important for the progress of the country's economy and for building the transparent, accountable governance architecture required for EU membership.
For the effective management of public resources in Albania, public financial systems are vital, especially in light of the ongoing reform agenda

and the goal to meet EU standards. The two main pillars examined — Public Accounting and Public Audit — serve complementary functions: systematic management of public revenues and expenditures in line with international standards, on the one hand, and independent oversight and reliability assurance of financial management on the other. Despite meaningful progress, challenges remain. Fragmented accounting practices, a low implementation rate for audit recommendations, inconsistent enforcement of commitment controls, high tax arrears, and limited independence of internal audit units are persistent weaknesses that must be addressed. The findings from the 420-respondent questionnaire closely align with the institutional assessments by SIGMA (2024), the World Bank (2025), and the IMF (2025), reinforcing the study's empirical conclusions. The interaction between accounting and auditing is generally seen by practitioners as enhancing transparency and accountability, but the coordination mechanisms between the two functions need strengthening through better information systems, standardized reporting, and more robust parliamentary follow-up. The Public Financial Management Strategy 2023-2030 and the National Reform Agenda 2024-2027 provide a strategic foundation for addressing these gaps. However, the OECD/SIGMA (2024) assessment that more than half of strategy performance indicators lack baseline values underscores the risk of reform without adequate measurement. Donors currently provide 93 percent of the additional funding for implementing the PFM Strategy, and this high dependency on external funding for reform costs can undermine long-term sustainability. Albania's rapid digital transformation — with over 95 percent of government services online via the E-Albania portal — offers a major opportunity to modernize financial reporting and audit processes. Recent Fintech developments in the public sector aim to encourage employees to exchange information in real time, which will increase the financial potential and opportunities for the development of further strategies. Plans to digitize public procurement using AI and to expand the Government Gateway platform for inter-agency data sharing are directly relevant to improving the transparency and efficiency of public financial governance.

Based on the research findings, the authors make the following recommendations for strengthening public financial governance in Albania. First, the recommendation implementation rate for ALSAI audit findings should be tracked through the newly established Parliamentary Sub-Committee for Public Sector Audit, with regular public reporting on compliance rates. Second, the transition to accrual-based accounting aligned with IPSAS should be accelerated, with dedicated technical assistance and annual progress benchmarks. Third, internal audit units should be granted greater functional independence, with audit plans approved at a level that insulates them from political interference. Fourth, annual professional development programs aligned with EU Internal Audit Standards should be institutionalized. Fifth, an integrated financial management information system (IFMIS) should be implemented across all budget entities to eliminate fragmentation in financial reporting. Sixth, international cooperation — particularly with EUROSAI member institutions and the EU's TAIEX programme — should be expanded to build auditor capacity and introduce best practices in performance and IT auditing.

A future development in accounting and auditing for the Albanian public sector focuses on increasing trust in experts, respecting values and ethical standards in daily financial administration and reporting. The demand for fairness and objectivity in the accounting and review processes has been expanded, emphasizing the respect of moral qualities and standards in daily administration. Auditing helps to protect against risk in the public sector; it provides advice and opinion to management on how to improve the performance of public entities, while public accounting provides the data infrastructure upon which good governance is built. At the current pace of EU convergence, it could take nearly a generation for Albania to reach half the EU average income (World Bank, 2025). Accelerating this convergence requires not only economic reform, but a fundamental strengthening of the accountability and financial governance architecture — of which public accounting and auditing are the indispensable foundation.

This study observed that the hypothesis testing shows different levels of support for public financial governance in Albania. Transparency was found to partially improve governance (H1), while auditing was fully supported in increasing accountability (H2). The interaction between accounting and auditing showed partial acceptance in improving trust and reducing mismanagement (H3).

The results derived from this analysis suggest that public auditing is essential for effective governance, despite problems with transparency and service quality. Recommendations for improvement include better coordination, digital integration and standardized procedures. Overall, improving the relationship between accounting and auditing is vital for increasing public trust and efficiency in Albania’s financial management.

## 6. REFERENCES

Albania, M. o. (2024). Public Finance Management Sectoral Strategy 2023-2030 and Action Plan 2023-2026. Council of Ministers Decision No. 390, dated 12 June 2024. Retrieved from https://financa.gov.al/

Bello, H. (2013). Audit committee's role in enhancing accountability of the Albanian Public Sector. European Journal

of Business and Economics, 8(4).

Blerina Sadiku & Majlinda Velcani. (2021). Political Culture as a Determinant Factor in the Performance of the Public Sector in Albania. Interdisciplinary

Journal of Research and Development, 8(1), 49.https://doi.org/10.56345/ijrdv8n104

Demi (Mosho) A., Dani M., & Kuci E. (2022). Money laundering, public accounting and auditing in Albania. Journal of Financial and Monetary Economics, 10(1), 54–60. https://ideas.repec.org/a/vls/rojfme/v10y2022i1p54-60.html

Demi (Mosho) A., & Puci J. (2016). The Structure of Financial Reporting and Accounting, Albanian Case. RSEP International Conferences on Social Issues and Economic Studies (pp. 103–110). Madrid, Spain. DOI: 10.19275/RSEPCONFERENCES016

Demi A. (2025). Monitoring and Risk Management Strategies in Public Policies and Programs: A Comprehensive Analysis. International Economic Relations and Business (pp. 79–85). Sofia, Bulgaria. DOI:10.37075/ISCWE.2024.07

Demi A., Farruku E. (2015). The framework of economic factors in Albania. Proceedings of the 5th International Conference on European Studies. Tirana, Albania.

Demi Mosho A. (2023). Socio-Economic Perspective in Albania, Objectives and Implementing Instruments. Interdisciplinary Journal of Research and Development, 10(1 S1), 228. https://doi.org/10.56345/ijrdv10n1s133

Demi, A., Xhaferri, S., Uku, S., Shahini, S., & Lushi, A. (2021). The impact of fiscal policies on Albanian economic growth: The case of value-added tax [Special issue]. Journal of Governance & Regulation, 10(4), 311–325. https://doi.org/10.22495/jgrv10i4siart11

European Commission. (2025). Albania 2025 Report (SWD 2025) accompanying the Communication on Enlargement Policy). Brussels: European Commission, Directorate-General for Neighbourhood and Enlargement Negotiations. Retrieved from https://enlargement.ec.europa.eu/

European Commission. (2024). Key Findings of the 2024 European Commission Report on Albania. Brussels: European Commission.

Grossi, G., et al. (2023). Changing the boundaries of public sector auditing. Journal of Public Budgeting, Accounting & Financial Management, 35(4), 417–430.

Hoxhaj, E. (2018). Public sector accounting reform in Albania, core challenge: Moving toward full accrual basis of accounting. The Macrotheme Review, 7(1), 49–63.

IFAC & CIPFA. International Framework for Good Governance in the Public Sector. Retrieved from https://www.ifac.org/

Imeraj, J., Hoda, N., Bica, K., Prenaj, V., & Gjoni, L. (2025). Exploring The Impact and Risks of Fintech Adoption on Income Inequality: A Global Cross-Sectional Study. Risk Governance & Control: Financial Markets & Institutions, 15(2).

IMF. (2025). Albania: 2025 Article IV Consultation — Staff Report. IMF Country Report No. 25/20. Washington D.C.: International Monetary Fund. Retrieved from https://www.imf.org/

INSTAT. (2022). Quarterly economic growth. www.instat.gov.al

ISSAI 100. (2019). Fundamental Principles of Public-Sector Auditing. International Standards of Supreme Audit Institutions. Retrieved from issai.org

Lireza, L. (2023). The Impact of Environmental Issues to Security in the Balkans. Interdisciplinary Journal of Research and Development, 10(1), 36. https://doi.org/10.56345/ijrdv10n106

OECD/SIGMA. (2024). Public Administration in Albania 2024 — Monitoring Report. Paris: OECD Publishing. Retrieved from https://www.oecd.org/

Puci, J., Demi, A., & Kadiu, A. (2023). Impact of macroeconomic variables on the construction sector. Corporate & Business Strategy Review, 4(1), 22–30. https://doi.org/10.22495/cbsrv4i1art2

Republic of Albania, Council of Ministers. (2024). National Reform Agenda 2024–2027. Decision No. 621, dated 10 October 2024. Retrieved from https://saspac.gov.al/en/reform-agenda-growth-plan/

Transparency International. (2025). Albania National Integrity System Assessment. Berlin: Transparency International. Retrieved from https://www.transparency.org/en/nis/countries/albania

World Bank Group. (2025). Albania Public Finance Review 2025. Washington D.C.: World Bank. Retrieved from https://www.worldbank.org/en/country/albania/publication/pfr-2025

Xhaferri, S., & Demi, A. (2015). Reassessment of fixed assets. Mediterranean Journal of Social Sciences, 6(5), 107–111.

Zenuni, B. R., & Miti, M. U. (2021). Accountancy Performance and Achievement (Accountancy Development Index), Case of Albania, European Journal of Economics and Business Studies. DOI: 10.26417/ejes.v9i1.p235-243

# VIRTUAL REALITY IN ROBOTIC CODING EDUCATION: AUTONOMOUS VEHICLE MODULE

***Ali Erduman*** [1] **,** ***Abdullah Kurtulmuş*** [2]

*1 Department of Electrical and Energy, Hendek Vocational School, Sakarya University of Applied Sciences, 54050 Hendek, Sakarya, Turkey*

*2 Graduate Education Institute, Department of Electrical and Electronics Engineering, Sakarya University of Applied Sciences, 54050 Serdivan, Sakarya, Turkey*

## Abstract

The importance of robotic system design in education is highlighted by the growing replacement of human labour by robotic and autonomous control systems. This change feeds the increased interest in curriculum integration of robotic coding. Mishandling the materials in robotic coding courses; however, might compromise safety and confidence and result in resource loss, financial pressure, and educational disappointments. It is recommended that training approaches move from virtual laboratories to real-world applications after thorough mastery to avoid such problems. An important tool for developing robotic coding applications is the training module based on virtual reality (VR), presented in this work. The module allows users to create autonomous vehicles by providing them with both visual and audio menus in a virtual reality environment. It includes parts such as sensors and control units. Users move from VR-based training to real-world applications with ease allowing for interactive learning that crosses both domains. This method is notable for improving student engagement, retention of information, and skill development. An analysis including 38 students, 20 in the experimental and 18 in the control groups, shows how effective robotic coding classes using the VR-based module are. Pupils demonstrate strengthened understanding, increased curiosity, and proactive participation in developing professional skills. Virtual reality workouts have been shown to enhance memory recall, professional competence, talents, and analytical ability. Additionally, this cutting-edge module expands engineering and educational frontiers, especially in the areas of autonomous vehicle design and external communication, giving instruction additional depth.

**Keywords:** Virtual Reality, Robotic Coding, Autonomous Vehicle, Robotic Education, Training Module.

## 1. INTRODUCTION

The century we live in is called the age of technology as it is a period of many advances in the field of technology (Kaya, 2006). The adaptation of technological developments to education has necessitated the emergence of new teaching methods and techniques. Education can be defined as the process of creating desired behaviors in the individual. Technology can be defined as the use of the knowledge and skills gained by the individual with the education he/she has received to make his/her life more comfortable (Aydoğan & Ersozlu, 2009). In this period, also called the age of technology, innovative technology applications such as robotic coding, virtual reality systems, and computers have been integrated into education programs, and they have taken their place in curricula. According to today's understanding of education, the transfer of application-based courses such as Robotic Coding education to students without practice or application is incomplete. Technological In today's educational understanding, according to the current lesson plan, the transfer of application-based courses such as Robotic Coding education to students without practice or application is incomplete. Due to the high cost of application materials, the inability to replicate them as much as they want emerges as an important problem in the learning process. Another important problem is that everyone is taught in the same educational model without taking into account the student's own competencies and performance. In this case, students compare themselves with their classmates and have a feeling of inadequacy, which leads to their dislike of the course.This study compares the structure of the traditional Robotic Coding course with its VR-enhanced counterpart. In the traditional format, the control group received instruction using standard textbooks and printed materials. In contrast, the experimental group engaged with a VR module that enabled them to design and operate a virtual autonomous vehicle interactively. A virtual reality training module was prepared to be used in robotics coding courses given in programs of vocational-technical education institutions such as informatics, electrical and electronics, mechatronics, etc. Following, the learning behaviors and attitudes of the students were analyzed.

Research has shown that in terms of the educational process in applied education, students who receive

*Corresponding author: Ali Erduman, alierduman@subu.edu.tr

education with educational materials complete their education 20% faster and more accurately than students who receive education with printed materials (Sirakaya, 2016). The innovative and easy-to-apply structure of the virtual reality environment, which is one of the subjects that students are most curious about in the age of technology, has increased interest in the lessons and made learning effective. Due to the impact of various stimuli on their perceptions, students who use VR technology in the classroom learn more effectively (Tepe, 2016).The practical parts of the courses held in schools involve a number of difficulties (Botan et al., 2023). These challenges include high costs, time-consuming repetition of training, difficult individualization and the lack of qualified training personnel. Maintaining the quality of education while minimizing costs is one of the main technical education research topics that academics are dealing with (Junior et al., 2013). Studies are being carried out with virtual reality to reduce the cost of technical education and improve the quality of education (McGovern et al., 2020).

In our age, robotic and autonomous control systems are replacing manpower (Heyer, 2010). Countries that want to advance in science and technology want to increase the interest of all their citizens in scientific technology (Ramachandran et al., 2021). In the studies conducted by academicians, it was generally concluded that the diversity of variables in robotic coding education is effective for student success (LeTendre & Gray, 2024). The decrease in robot kit costs, and the increase in interest and accessibility have affected this situation (Yolcu & Demirer, 2017) . Due to the sustainability of development in the world and the fact that all digital technologies are programmable, coding education is becoming increasingly important in the education of trained manpower (Karataş,2021). It was found that while students' effective use of computers increased, their anxiety decreased, and their self-efficacy perceptions such as cognitive thinking and information processing increased (Ramazanoğlu, 2021). Since 1997, the use of digital devices as educational tools and their impact on digital learning has been increasing (Reiser, 2001). In recent years, virtual reality (VR) technologies have been actively used in education and training environments with various applications (Jensen & Konradsen, 2018).

According to a Portuguese study, using VR in the classroom has a good impact on students' education and offers hope for the future of education (de Moraes Rossetto et al., 2023). Companies such as HTC Vive, Samsung Gear VR, or Oculus Rift enable the creation of a real-world perception in a non-physical virtual world with virtual images created with head-mounted displays (Smith & Bridle, 2018). Exercises with VR glasses are said to help students learn and retain information better (Radianti et al., 2020). VR has been a challenge in terms of developing professional skills, professional knowledge, and sound judgment (Gallagher et al., 2005). A review of 86 scholarly works on game-based learning in electrical engineering from 2001 to 2021 has found that 56% of them used virtual reality (Giraldo et al., 2024). To be used in engineering education in China, a VR-based training module has been created to greatly enhance students' theoretical knowledge, practical skills, and innovative abilities to make graduates highly sought-after by employers (Wang & Liu, 2024) . The Malaysian education system is incorporating metaverse and virtual reality studies into mechanical engineering education. This training module aims to meet the current demands in the field of mechanical engineering in Education 5.0. and adapt to the transformation of the industry (Sidhu et al., 2024) . A VR learning system was designed to offer medical students the chance to practice complex medical skills and solve clinical problems with immediate feedback, and it was found that their medical skills were significantly better than the control group, showing higher learning motivation and self-efficacy (Lin et al., 2023). At this point, researchers, educators, and organizations are following the increasing development of VR technologies to add an extra dimension to the teaching and learning environment (Yu & Xu, 2022). Compared to traditional instruction, the use of wearable hybrid AR/VR learning materials significantly enhanced students' situational interest, emotional engagement, and learning performance in science education (Sun et al., 2023).

In robotics education, it is very important to use laboratories or virtual laboratories that allow the application of the acquired knowledge and the observation of real errors or problems that can occur in a real system but not in a simulation (Craifaleanu & Craifaleanu, 2022). To analyze the features of VR at a more abstract level, B. Chavez et al. implemented the virtual laboratory in a virtual laboratory to simulate the robot before real experiments were performed (Chavez & Bayona, 2018). Gabajová et al. designed a virtual workplace application in the Unity 3D game engine by modelling the real environment in 3D so that unnecessary waste of resources can be avoided with virtual workplaces and the same work can be carried out (Gabajová et al., 2021). R. Kumar et al. showed that students retained more information and were better able to apply what they learned after participating in VR exercises (Kumar, 2014). In a meta-analysis evaluating the effectiveness of simulations, a three-dimensional world was created and designed as an easy, understandable, and fun teaching environment for the realization of learning with virtual reality in a study of 821 nursing students (Tuna & Öskan, 2022). It demonstrated the importance of developing content for the VR environment to achieve effective learning by leveraging the unique advantages of technological advances in VR science lab simulations (Makransky

et al., 2019). There is a significant difference between competency development in a VR classroom and a regular classroom with a similar curriculum (Lee & Shvetsova, 2019). W. Natephra et al. demonstrated that building information modelling with the Internet of Things and augmented reality (AR) virtual reality can be controlled and observed with sensors and microcontroller cards (Natephra & Motamedi, 2019). Küçükkara et al. created interactive objects on the Unity platform and provided an immersive experience. In addition, thanks to the Meta Avatar plugin, users can create their own avatars and edit unique platforms with this avatar (Küçükkara et al., 2023). With plug-ins such as the Mixed Reality Toolkit (MRTK), touch-based interactions have been created through hand tracking, making it easier to experience objects (Ho et al., 2022).

In the current era, educators can gamify the learning process by making it fun and meaningful in order to integrate technology into the learning process in terms of content and pedagogy (Kapp, 2012). It made it possible for students to feel the sensation of the virtual world with their visual and auditory sensory organs by using virtual reality technology (Wu et al., 2023). Agustini et al. concluded that there was a 31.62% increase in student scores in the pre-test and post-test analyses of the narration with the game they created (Agustini et al., 2023). VR has positive effects on the planning, testing, thinking, and appreciation phases of design processes (Slezaka et al., 2023). Creative observation and innovative application development have an effective and significant role in improving the performance of processes and outcomes (Chang et al., 2023). VR was found to be more effective in art history education in terms of students' learning achievement, metacognitive awareness, self-regulation and competence (Wu et al., 2023).

In the virtual laboratory, the user has the chance to practice trial and error and practice over and over again. There is no chance of job hazard or breakdown during these stages. It also simulates students developing a simple robot model together in the same space. In the virtual laboratory study, the first software debugging was performed before the students were tested on the real robot in a safe and secure way, and they fully agreed that the laboratory was necessary with the questionnaires applied to the students (Chaos et al., 2013). Luiz A. Junior et al. showed that it is more feasible to create an educational low-cost robot kit for students (Junior et al., 2013). In Turan's study, a mobile system design was provided that will automatically follow the line if the person wants the battery-powered disabled vehicle and the obstacle in front of the vehicle, calculating the environmental conditions and orientations according to the situations that will occur. Also, if desired, movement control can be provided with the help of other people (Turan, 2017). In their study, S. Kaya et al. combined robotic coding training and the "Internet of Things" in digital transformation and prepared an experimental set that can control the external unit with the relay on the Arduino card, which can provide wireless communication, and various sensor connections can be made with an RJ45 connection (Kaya & Samtaş, 2021). The advantages and disadvantages of real autonomous driving were calculated and studies were conducted and analyzed by utilizing different motion planning techniques (Nennioğlu & Köroğlu, 2018). With the diversification of the sensors to be added to the educational robot and the "Internet of Things" additions, it is ensured that it evolves into situations that will provide the beginning of industrial breakthroughs that will lay the foundations of electric vehicles.

The third section of the article describes the creation of autonomous vehicles and virtual laboratories with virtual reality. the fourth section presents the findings obtained and the final section provides conclusions.

## 2. MATERIAL AND METHOD

In this study, the formation and execution architecture of our autonomous training vehicle system, which is given in real and virtual environments in a multi-layered manner, is graphically given in Figure 1. As stated in the figure, 3D modelling of real robot parts was created and designed in a VR environment thanks to the game engine. Thanks to the plug-ins integrated into the game engine, an interactive training module was designed and an autonomous robotic vehicle was assembled. After the design was completed, it was started to be designed in the real environment. The vehicle, which was assembled in the real environment, was connected to the virtual environment and operated.

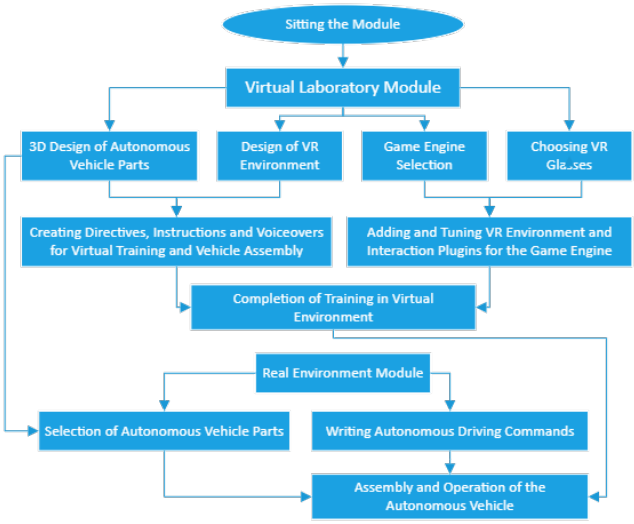



**Figure 1:** *Integrated Autonomous Vehicle Training Module: Real and VR Environment Schematic*

In the study conducted within the scope of this

research, an application was developed for vocational high school students, which includes the assembly of the training robot in the robotic coding training program. In the application, the explanations of robot parts that are difficult to understand in robotic coding education were included in virtual education in audio and visual form. Audio recordings taken from students of the appropriate age for the training group were converted into Waveform Audio File (.wav) format and added to the game engine. In visualization processes, objects and environments were drawn with 3D modelling tools. The objects used in the menu were photographed, and the images in their backgrounds were deleted and converted to Portable Network Graphics (.png) format with a transparent background and used in the game engine. Virtual reality user-friendly interface sample material information and visual interface examples are given in Figure 2. These visual interfaces provide detailed information about how and for what purpose the materials used in the working environment will be used. Students can learn the materials used in robotic coding in the virtual environment by reading from the introduction and feature menus and assembling the autonomous vehicle with audio and visual instructions.

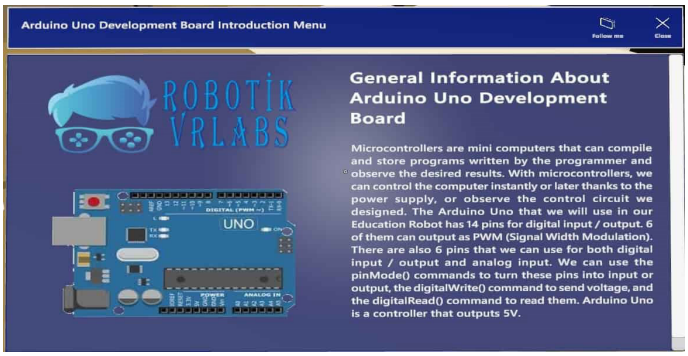



**Figure 2:** *Virtual reality user-friendly interface and sample material information visual interface*

The study was carried out in two main parts. In the first part, the virtual reality training module was designed and produced. 3D modelling of the materials and laboratory to be used in robotic coding education was created. Interactions were added to the modelled objects with Meta Quest 2 plugins and coding to create the application for students in a VR environment. In the second stage, the materials required for the realization of the autonomous vehicle to be developed were selected and the assembly of the real autonomous vehicle was completed. If the coding was done in a way to provide autonomous driving, and the virtual training was approved by the system, the training robot would be operated from the virtual environment and it would be enabled to work in the real environment.

### *2.1 Robotic Coding Module in Virtual Laboratory (RCMVL)*

Blender software was used to create the drawings of the virtual laboratory, and Unity game engine software was used to interact with the created objects. The object and the laboratory, which were photographed in order to take samples from the real world, were 3D modelled with the Blender program and then converted into a FilmBox (fbx) format that can be used with Unity. In Figure 3, the real laboratory, on the left side, and the images of the virtual laboratory created based on the real classroom, on the right side, are given from different angles.

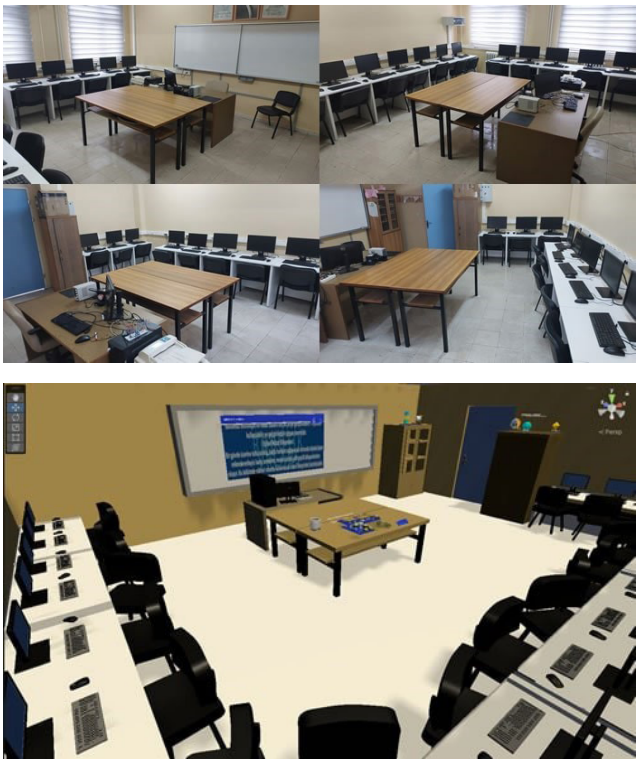


**Figure 3:** *Real robotics lab (Up) and 3D modelled robotics lab (Down).*

Virtual reality glasses (Meta Quest 2) were used to test and use the training module, which was prepared with the Unity game engine and created in the virtual laboratory, in VR environment. The information received from the Meta Quest 2 sensors was processed with MRTK, which is used to provide various interactions in the virtual reality environment. Meta Quest 2's features such as touching, moving and interacting with objects in the virtual environment with our own hands, without the need for a controller tool or any gloves, were added to the virtual reality module with MRTK plugins.

At the end of the training in the virtual reality environment, data was transferred from the virtual environment to the real world, and Bluetooth serial communication system was used for autonomous

vehicle control through which real vehicle was controlled. In the preparation of all these processes, Microsoft Visual Studio Community program and C# language were used to make laboratory modeling interactive with the Internet of Things.

### *2.2 Creating RCMVL*

The algorithm used to create RCMVL is given in Figure 4. When the training first starts, the instructions describing the training and the instructions that the user must follow are given audibly and visually. In order to show and guide the user the shape of the autonomous vehicle, 3D objects were created in the virtual environment and the stationary robot was animated until the user touched it. When the user picks up an object, the object menu and audio narration are opened and information is provided. When the object was placed at the right point and in the right order, the position, size and orientation were adjusted and placed. Thus, the user was helped to complete the assembly of the autonomous vehicle correctly and complete the training. When the assembly was completed, a warning was given to start the vehicle. If the user presses the start button of the vehicle, the autonomous vehicle is contacted via serial communication and the vehicle in the real environment is started.

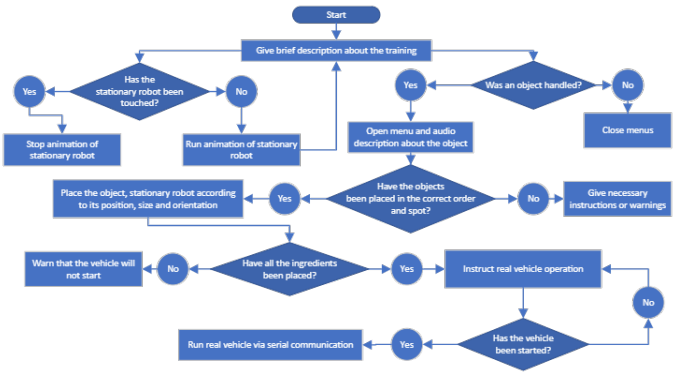



**Figure 4**. *SLRKM creation flowchart*

The stationary vehicle was created as a transparent material as shown in Figure 5, and all its parts were made transparent. When the training was first started, a color transition animation from pink to green was created transparently to attract the attention of the students until the robot was clicked. The animation was automatically paused when the student touched the first object.

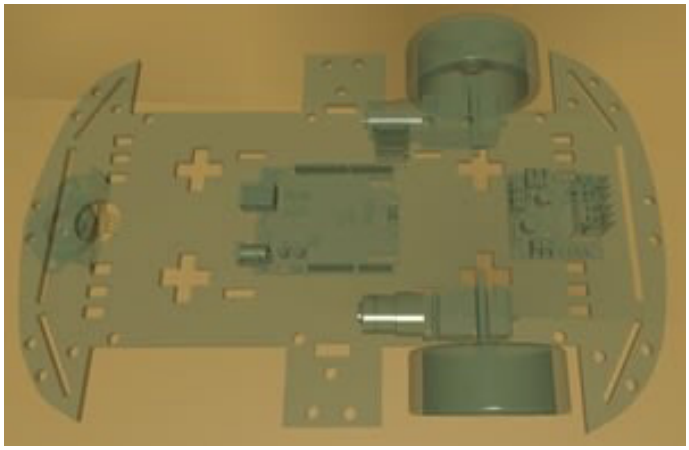


**Figure 5:** *Example transparent training robot vehicle model view in virtual lab.*

Some components have been added to the objects that make up the stationary robot in order to provide interaction when touched by other objects. Physical properties are given with Rigidbody, which is one of them. The position and rotation angles were locked so that the Rigidbody could not be moved from the Rigidbody settings. Also, the touch area of influence was determined with the Collider setting, and it was made clear which object was touched by giving the Tag name. Thus, when the user places the correct part at the point where it should be placed, the object in the student's hand is placed on the fixed robot like a magnet with the following lines of code. The function commands "gameObject.transform.position = degerGirildi.position" to change the position of the object with the position of the fixed robot, "gameObject.transform.rotation = degerGirildi.rotation" to change its angle and "gameObject.transform.localScale = degerGirildi.localScale" to change its size were used.

To enable users to be able to pick up the objects, enlarge and resize them, and to activate the voice and commands that can be added, some components were selected and used among the components that come with the MRTK plugin. Some of the components that come with the MRTK plugin are as follows: Constraint Manager, Near Interaction Grabbable, Cursor Context Object Manipulator. The added Object Manipulator component and its settings menu are shown in Figure 6. Thanks to the manipulator component given in Figure 6, the interactive states of the objects used in the virtual laboratory can be changed, activated, and deactivated.

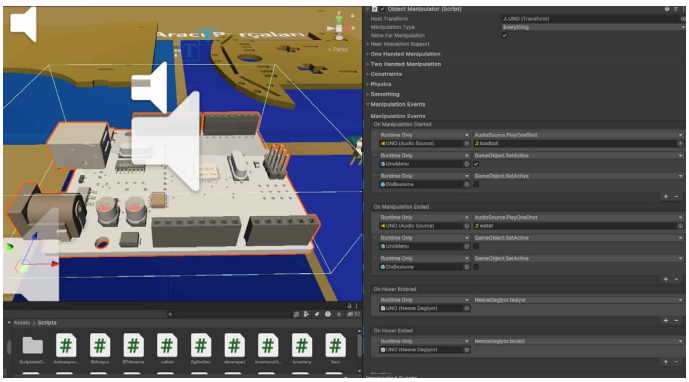


**Figure 6:** *The MRTK in Object manipulator function selection interface.*

At the moment when the student carries the objects in his/her hand, a menu with the introduction of the robot parts was prepared with a pop-up menu as given in Figure 2, and the students were given visual training that was instantly active in the simulation. In the pop-up training menu, the features of being active while the student is carrying the material in his/her hand, choosing to follow the training menu with the other hand at the distance of reading glasses, closing the menu and opening the texts that can remain at the bottom with the scroll bar have been added.

To enable virtual reality environment to attract the user's attention and to have an interactive structure, sound effects and voice instructions were included in the training. In this way, the instructions that the student will fulfill are given from the very beginning of the game and how to continue the simulation in the game is given in a hierarchy. Audio Source components were added to activate the sound files in the objects and to give audio warnings. When the student picks up the object, the activation of the objects and the control of the sound files we want to play are controlled with the code file named object touches. The audio files were added and the sound was played by changing the Audio Clip file of the active object in the coding.

In the in-practice evaluation of the training, a different kind of tool was added to each category and introduced and the correct choice was expected. The selected product was asked to correctly assemble the virtual robot by following the audio and visual instructions in the training. The completion of our virtual reality training was checked with a code in which we observed the placement of each of the fixed robot parts. When all the parts were in place, an audible warning was given to indicate the completion of the training and instructions on how to operate the vehicle in the real environment.

With the menu shown in Figure 7 at the bottom right of the training, the buttons for restarting the training, activating the hand beam, glove view, hand joint points and running the training vehicle in the real environment were added. When the training was completed, the button to start the training tool in the real environment was designed in different colors with an animation to attract attention. The system is activated to run the real tool. If the user pressed the button to run the training tool in the real environment without completing the training, the user was instructed to complete the training first.

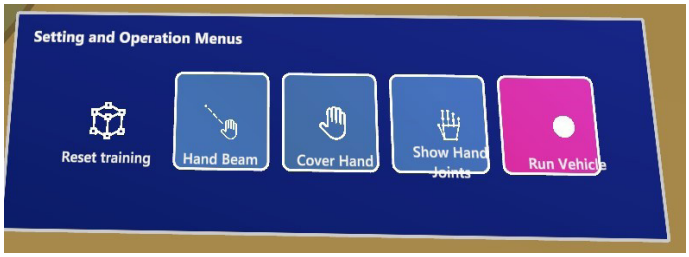



**Figure 7:** *Setup and operation control setting menu.*

### *2.3 Creating the Autonomous Robot*

Students who reached sufficient competence in the autonomous vehicle created in the virtual environment were asked to produce the physical vehicle. To prevent students from escaping by rote, different products such as motors, wheels, sensors, etc. were placed on the menu, and they were asked to choose the right products. In this context, students were asked to assemble the robot according to the workflow sequence in the virtual environment. The image of the autonomous vehicle whose production was completed is given in Figure 8.

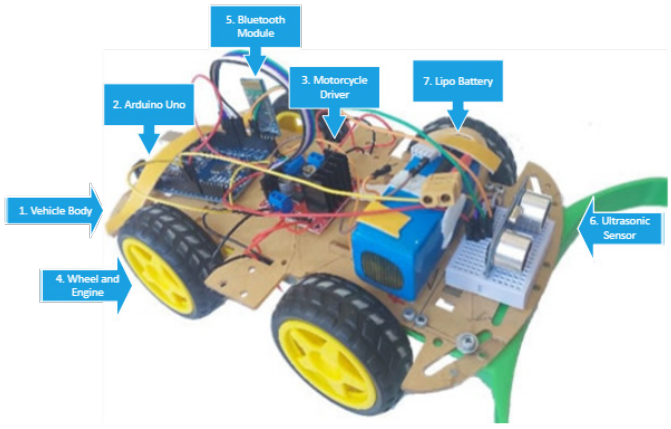



**Figure 8:** *View of the assembled training robot..*

The actual parts used in the autonomous vehicle and their descriptions are listed in Table 1

| No | Descriptions |
|---|---|
| 1 | The body created with plexiglass was used in the training robot. |
| 2 | Arduino UNO microcontroller card is a mini-computer that can be controlled with the coding written or the designed control circuits can be observed. |
| 3 | L298N motor driver, which can control two motors at the same time and can also control the speed depending on the frequency. |
| 4 | Easy to assemble and economical side geared plastic motor (right) and easy to assemble wheel. |
| 5 | HC-06 Bluetooth technology was used to communicate with the vehicle in the real environment when the training created in the virtual environment was finished. |
| 6 | HC-SR04 ultrasonic distance sensor is used to orient the autonomous vehicle according to the object in front of it. |
| 7 | Lipo battery is used for long-term energy supply and easy charging features. |

**Table 1:** *Training module components*

While acceleration is achieved by rotating the two wheels at the same speed, angular rotation is realized by rotating the wheels at different speeds for the rotation function. In the back part of our educational robot, 2 wheels are used for propulsion, while in the front part, wheels that can rotate in all directions are used to ensure the balance of the robot. Thanks to the ultrasonic distance sensor, the vehicle is given an autonomous driving capability.

Equation 1 is used to convert the distance of the rotating signal from a unit of time to a unit of distance in cm

$$L = \frac{t * 0{,}343}{2}$$

Where L is the distance unit in meters, t is the time in seconds.

Thus, when the virtual reality VR training is completed, the flow chart for the autonomous movement of the real vehicle operated from the virtual environment is given in Figure 9. When the training created in the virtual environment was completed, coding was done to start the vehicle with serial communication from the VR environment with the help of Bluetooth module. The autonomous vehicle will wait without starting until the vehicle in the virtual environment is

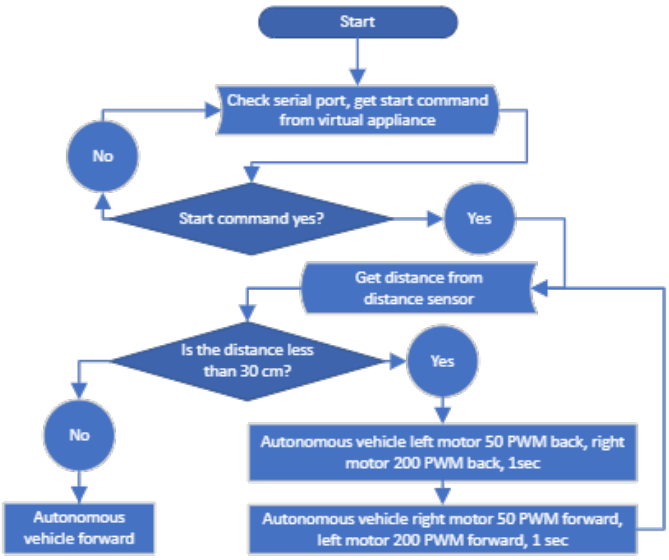



**Figure 9**: *Program flowchart of an autonomous vehicle.*

When the start command comes, it will go forward as long as the data from the distance sensor is not less than 30 cm. If the distance is less than 30 cm, control commands are sent to the motor driver to give the vehicle driving capabilities to avoid the obstacle. The algorithmic maneuver driving situation that the autonomous vehicle will follow in case an object appears in front of it is shown in Figure 10 below.

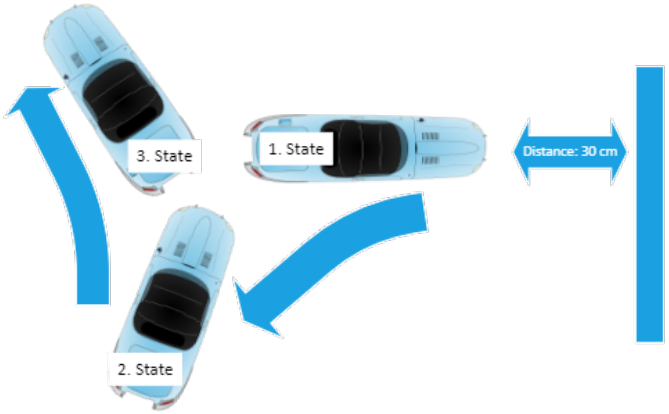



**Figure 10:** *Driving state diagram of an autonomous vehicle.*

Arduino IDE compiler application was used for programming our autonomous vehicle. With Arduino IDE, the microcontroller was programmed using C++ language and the program was compiled and loaded.

### *2.4 Attitude scale towards VR usage*

In this study, the attitude scale given in Table 2 and developed by F. G. Karaoglan et al. was used

(Karaoglan-Yilmaz et al., 2023). In order to measure the attitudes of Information Technologies Department students of vocational high schools towards their educational programmes, a simple tool has been developed and tested for the attitude scale. The use of virtual reality in Information Technologies department Robotic Coding class has been evaluated by nine attitude scale and each of these scales has been evaluated by Likert scale.

As a result of the tests, Cronbach's alpha coefficient value and test-retest reliability were taken into consideration. it was confirmed that each item had discriminative power with significant t-values and high total variance. The study group of our research was determined as 10th grade students who continue their education in the field of Information Technologies at Vocational and Technical Anatolian High School affiliated to the Ministry of National Education.

| Item | Questionnaire |
|---|---|
| 1 | Virtual reality increases my study efficiency |
| 2 | Using virtual reality for learning boosts my productivity |
| 3 | Using virtual reality increases my learning efficiency |
| 4 | Using virtual reality for learning purposes helps my professional development |
| 5 | I find using virtual reality beneficial for learning purposes |
| 6 | I think it's a good idea to use virtual reality for learning purpose |
| 7 | I will encourage people around me to use virtual reality for learning purposes |
| 8 | Using virtual reality for learning is compatible with my professional perspective |
| 9 | Using virtual reality for learning helps solve problems I encounter in learning |

**Table 2:** *Attitude measurement questionnaire for the use of virtual reality in robotic coding education (Karaoglan-Yilmaz et al., 2023).*

## 3. DESCRIPTION OF RESEARCH USING THE VR SYSTEM AND PRESENTATION OF THE RESULTS

Our research was planned on a quantitative basis and conducted in two stages. In the first stage, we measured the desired behaviors of students in Robotic Coding education and the targeted changes after the training they received using the test method. In the second stage, we assessed their attitudes and judgments towards virtual reality and their views on virtual education. We interpreted their contributions to education by analyzing numerical data on their tendencies towards the use of virtual reality in education. In the 2023 - 2024 academic year, the students, who are planned to make educational robots with Robotic Coding Training, have received training to improve their theoretical and assembly practices on arduino and robot parts. Our study group was divided into an experimental group of 20 students and a control group of 18 students. The preparation of the study environment and its application to the students is given in Figure 11

**Figure 11:** *Test and Attitude Scales Applied During VR Robotics Education Implementation.*

Within the scope of this study, the educational robots and assembly system documents prepared by the Ministry of National Education for 10th-grade students in the field of Information Technologies in

Vocational and Technical Anatolian High Schools were used as reference for the Robotic Coding course material and assessment and evaluation questions related to the subject (Altın et al., 2024).

In the data in Table 3, the mean pretest scores of the experimental and control groups were 3.45 and 3.78, respectively. There was no statistically significant difference between the control and experimental group students in the independent sample t-test results of the pretests applied to the students, t=-0.756, p>.05.

| | Group | N | Average | t | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Pretest | 1 | 20 | 3,45 | -,756 | ,454 |
| | 2 | 18 | 3,78 | -,750 | ,458 |

**Table 3**: *Statistics of pretest mean scores of control and experimental groups*

In the results of the independent sample t-test of the difference between the pretest and posttest applied to the students, a statistically significant difference was found between the control and experimental group students, t=10.975, p<0.05.

The 20 participants' answers to each question were scored as 1 (correct) and 0 (incorrect). In Figure 12, the answers given for each question are summarised and the results of the application as pretest and posttest are interpreted graphically. The average of the pretest and posttest answers given by the experimental group students to the assessment and evaluation questions increased from 6.45 to 14.27, respectively. This shows that there is a significant change in the learning levels of the students.

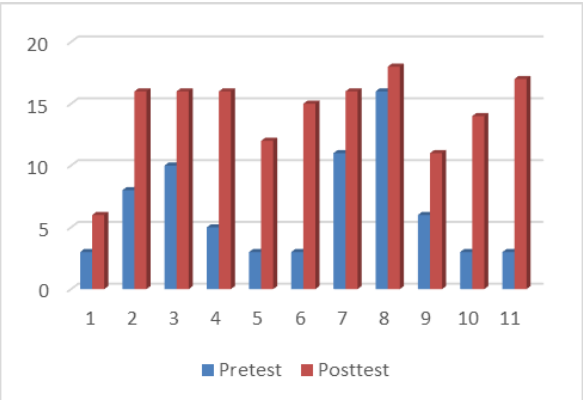



**Figure 12**: *Measurement and evaluation pretest-posttest difference graph*

### *3.2 Educational Virtual Reality Attitude Scale Questions and Implementation*

In this study, in order to determine the change in attitudes towards the use of VR technology for educational purposes the test prepared by F. G. Karaoglan. et al. was used (Karaoglan-Yilmaz et al., 2023). As a result of the tests, Cronbach's alpha coefficient value of 0.92 and test-retest reliability of 0.94 were found, and as a result of the reviews and revisions, the attitude scale consisting of nine attitude items and a five-point Likert structure was prepared in Turkish and English.

In the second stage, The educational robotic module developed in the artficial reality which serves as a tool in the training of SLKRM users has been used in artificial labarotary. In the second stage, the scale reults on the users' tendency to use the modüle and the data has been presented in this section. A pre-test was applied in both groups. In Table 4, the mean scores of the pre-test attitude scale applied to our experimental and control groups were 28.55 and 28.28, respectively. There was no statistically significant difference between the control and experimental group students in the independent sample t-test results of the pre-test scale applied to the students, t=-0.158, p>.05.

| | Group | N | Average | t | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Pretest | 1 | 20 | 28,55 | ,158 | ,876 |
| | 2 | 18 | 28,28 | ,156 | ,877 |

**Table 4**: *Statistics of the mean scores of the control and experimental groups on the pre-attitude scale.*

After the lectures were over, the final attitude scale was applied to both groups. The difference between the pre and post-attitude scale applied to our experimental and control groups was found to be 9.35 and 1.39, respectively. In the results of the independent sample t-test tests of the difference between the pre and post-attitude scale applied to the students, a statistically significant difference was found between the control and experimental group students, t=8.477, p<0.05. The answers of 20 participants to each question were numbered from 1 to 5 on a Likert scale and the average of the pre and post-attitude scale is graphically interpreted in Figure 13. The average of the answers given by the experimental group students to the attitude scale increased from 3.36 to 4.02.

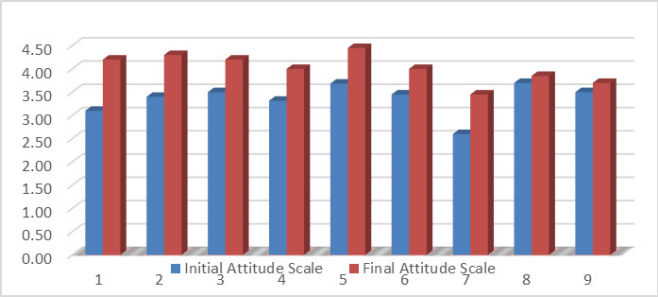



**Figure 13**: *Difference graph between pre and post attitude scale of the experimental group*

In the health and safety warnings of virtual reality goggles, it is written that the maximum usage time should be 30 minutes and 10-minute breaks should be (Agustini et al., 2023). The time spent by the experimental group students with the virtual reality goggles during the application was measured. In the measurements, it was found that the students who used the first application finished the application in 7 minutes at most, and the time of the students who watched their friends' application on the screen while their friends were doing it decreased to approximately 5 minutes. When the application was reapplied to a group of students one week later, it was found that the students who used the application for the second time completed the robotic coding training in the virtual laboratory with an average of 3 minutes. Since the maximum usage time was 7 minutes in the measurement results, the virtual reality glasses had a value below the health and safety warnings.

The main contributions to the literature within the scope of this study are given below.

1. A hybrid course module for designing an autonomous vehicle with audio, visual, and interactive VR tools was created for the Robotics Coding course.

2. By modeling the laboratories that students are currently using, a visual integrity between the virtual and real environment was provided.

3. The autonomous vehicle module, which is a subject of the robotics coding course, was created virtually in the virtual laboratory and in case of successful completion of the module, it was enabled to work interactively with the real world.

4. The consistency of the study was analyzed with test groups and its effects on the educational environment were examined. As a result of the analysis, it was determined that the use of virtual reality has positive effects on the educational environment.

## 4. SUMMARY AND CONCLUSIONS

Within the scope of this study, a course material was prepared to help students practice in robotic coding courses with the robotic coding training module in the virtual laboratory created using VR technology. In this context, the training robot in the Robotic Coding Course Material was built in a real environment by gaining autonomous driving capability. All parts of the autonomous robot and the robotic coding laboratory used for the lessons were 3D-modelled with the help of the Blender 3D drawing program. Within the scope of this research, an experimental group of 38 students, 20 experimental and 18 control group students from the 10th-grade Information Technologies Department, was studied. According to the results of the attitude scale before the application, the results of the independent sample t-test tests were $t=-0.158$, $p>0.05$. In line with the answers given by the participants to the attitude scale questions, it was revealed that using virtual reality will increase their working efficiency, their productivity and improve their professional performance.

As a result, the prepared module eliminated the problems such as working with insufficient materials and material losses that may occur as a result of experiments in creating educational robots for students taking robotic coding courses. In the created virtual environment, it was possible for each student to reinforce the subject, and students were offered the chance to repeat complex robotic coding skills over and over again. In the learning process, it increased the interest in the robotic coding course and contributed more actively to the acquisition of professional knowledge. The VR environment exercises that the students did helped them learn and retain the information better. With this innovative module, an extra dimension was added to the learning and teaching environment and autonomous robotic vehicle training and communication was provided with the VR environment.

In the following studies, it is planned to develop training modules by expanding the training modules in the virtual laboratory, especially in informatics and electrical-electronics engineering.

**Acknowledgments.** This work was supported by Scientific Research Projects Coordination Unit of Sakarya University of Applied Sciences. Project Number: 198-2024

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## 5. REFERENCES

Agustini, K., Putrama, I. M., Wahyuni, D. S., & Mertayasa, I. N. E. (2023). Applying gamification technique and virtual reality for prehistoric learning toward the metaverse. International Journal of Information and Education Technology, 13(2), 247–256. https://doi.org/10.18178/ijiet.2023.13.2.1802

Altın, A., Akpınar, E., & Karayiğit, H. (2024). Bilişim teknolojileri alanı robotik ve kodlama. Millî Eğitim Bakanlığı Yayınları.

Aydoğan, I., & Ersozlu, Z. (2009). Introduction to education science. https://www.researchgate.net/publication/319262399

Botan, V., Williams, N., Law, G. R., & Siriwardena, A. N. (2023). The effect of specific learning difficulties on general practice written and clinical assessments. Medical Education, 57(6), 548–555. https://doi.org/10.1111/medu.15008

Chavez, B., & Bayona, S. (2018). Virtual reality in the learning process. In Advances in Intelligent Systems and Computing (pp. 1345–1356). https://doi.org/10.1007/978-3-319-77712-2_129

Chaos, D., Chacón, J., Lopez-Orozco, J. A., & Dormido, S. (2013). Virtual and remote robotic laboratory using EJS, MATLAB and LabVIEW. Sensors, 13(2), 2595–2612. https://doi.org/10.3390/s130202595

Chang, Y.-S., Chou, C.-H., Chuang, M.-J., Li, W.-H., & Tsai, I.-F. (2023). Effects of virtual reality on creative design performance and creative experiential learning. Interactive Learning Environments, 31(2), 1142–1157. https://doi.org/10.1080/10494820.2020.1821717

Craifaleanu, A., & Craifaleanu, I. (2022). A co-creation experiment for virtual laboratories of mechanics in engineering education. Computer Applications in Engineering Education, 30(4), 991–1008. https://doi.org/10.1002/cae.22498

de Moraes Rossetto, A. G., Martins, T. C., Silva, L. A., Leithardt, D. R. F., Bermejo-Gil, B. M., & Leithardt, V. R. Q. (2023). An analysis of the use of augmented reality and virtual reality as educational resources. Computer Applications in Engineering Education, 31(6), 1761–1775.

Gabajová, G., Krajčovič, M., Matys, M., Furmannová,B., & Burganová, N. (2021). Designing virtual workplace using Unity 3D game engine. Acta Tecnología, 7(1), 35–39. https://doi.org/10.22306/atec.v7i1.101

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., ... & Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Annals of surgery, 241(2), 364-372.

Giraldo, J. S., Kok, K., & Paterakis, N. G. (2024). A structured review on game-based learning applied to electrical power and energy engineering. Computer Applications in Engineering Education, 32(1), e22686.

Heyer, C. (2010). Human-robot interaction and future industrial robotics applications. In 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 4749–4754). IEEE. https://doi.org/10.1109/IROS.2010.5651294

Ho, S., Liu, P., Palombo, D. J., Handy, T. C., & Krebs, C. (2022). The role of spatial ability in mixed reality learning with the HoloLens. Anatomical Sciences Education, 15(6), 1074–1085. https://doi.org/10.1002/ase.2146

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies, 23(4), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Junior, L. A., Neto, O. T., Hernandez, M. F., Martins, P. S., Roger, L. L., & Guerra, F. A. (2013). A low-cost and simple Arduino-based educational robotics kit. https://www.researchgate.net/publication/276849083_A_Low-Cost_and_Simple_Arduino-Based_Educational_Robotics_Kit

Kapp, K. M. (2012). The gamification of learning and instruction: Game-based methods and strategies for training and education. Wiley. https://www.wiley.com/WileyCDA/Section/id-818130.html

Karaoglan-Yilmaz, F. G., Yilmaz, R., Zhang, K., & Ustun, A. B. (2023). Development of educational virtual reality attitude scale: A validity and reliability study. Virtual Reality, 27(3), 1875–1885. https://doi.org/10.1007/s10055-023-00778-z

Karataş, H. (2021). 21st century skills in robotics and coding training Turkey and the world. Education and Society in the 21st Century, 10(30), 693–729.

Kaya, S., & Samtaş, G. (2021). Orta ve ileri düzey robotik kodlama eğitimleri için internet odaklı sensör kartı tasarım ve imalatı. Manufacturing Technologies and Applications, 2(3), 78–88.

Kaya, Z. (2006). Instructional technologies and material development. Pegem A Publishing.

Kumar, R. (2014). Araştırma yöntemleri: Yeni başlayanlar için adım adım araştırma rehberi (Ö. Çokluk, Çev. Ed.). Edge Akademi.

Küçükkara, M. Y., Özacar, K., & Ortakci, Y. (2023). Architecture students' experience of surveying in Safranbolu tannery in virtual reality environment. Fırat Üniversitesi Mühendislik Bilimleri Dergisi. https://doi.org/10.35234/fumbd.1322782

Lee, J. H., & Shvetsova, O. A. (2019). The impact of VR application on student competency development: A comparative study of regular and VR engineering classes with similar competency scope. Sustainability, 11(8). https://doi.org/10.3390/su11082221

LeTendre, G. K., & Gray, R. (2024). Social robots in a project-based learning environment: Adolescent understanding of robot–human interactions. Journal of Computer Assisted Learning, 40(1), 192–204. https://doi.org/10.1111/jcal.12872

Lin, H., Hwang, G., Chou, K., & Tsai, C. (2023). Fostering complex professional skills with interactive simulation technology: A virtual reality-based flipped learning approach. British Journal of Educational Technology, 54(2), 622–641. https://doi.org/10.1111/bjet.13268

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

McGovern, E., Moreira, G., & Luna-Nevarez, C. (2020). An application of virtual reality in education: Can this technology

enhance the quality of students' learning experience? Journal of Education for Business, 95(7), 490–496. https://doi.org/10.1080/08832323.2019.1703096

Natephra, W., & Motamedi, A. (2019). Live data visualization of IoT sensors using augmented reality (AR) and BIM. In Proceedings of the 36th International Symposium on Automation and Robotics in Construction (pp. 632–638). https://doi.org/10.22260/isarc2019/0084

Nennioğlu, A. K., & Köroğlu, T. (2018). Otonom araçlarda hareket planlaması. Artıbilim: Adana Bilim ve Teknoloji Üniversitesi Fen Bilimleri Dergisi, 1(2), 20-29. https://izlik.org/JA62FW59JG

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education, 147, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. Earth and Space Science, 8(5). https://doi.org/10.1029/2020EA001562

Ramazanoğlu, M. (2021). The effect of robotic coding applications on middle school students' attitudes towards computer and their self-efficacy perceptions towards computational thinking skills. Türkiye Sosyal Araştırmalar Dergisi, 25(1), 163–174.

Reiser, R. A. (2001). A history of instructional design and technology: Part I: A history of instructional media. Educational Technology Research and Development, 49(1), 53–64. https://doi.org/10.1007/BF02504506

Sidhu, M. S., Mousakhani, S., Lee, C. K., & Sidhu, K. K. (2024). Educational impact of metaverse learning environment for engineering mechanics dynamics. Computer Applications in Engineering Education, e22772.

Sirakaya, M. (2016). Use of augmented reality in applied education: Motherboard assembly [Master's thesis, Kırşehir Eğitim Fakültesi].

Slezaka, R. J., Keren, N., Gilbert, S. B., Harvey, M. E., Ryan, S. J., & Wiley, A. J. (2023). Examining virtual reality as a platform for developing mental models of industrial systems. Journal of Computer Assisted Learning, 39(1), 113–124. https://doi.org/10.1111/jcal.12731

Smith, R., & Bridle, O. (2018). Using virtual reality to create real world collaborations. Proceedings of the IATUL Conferences, Paper 5. https://docs.lib.purdue.edu/iatul/2018/collaboration/5

Sun, J. C.-Y., Ye, S.-L., Yu, S.-J., & Chiu, T. K. F. (2023). Effects of wearable hybrid AR/VR learning material on high school students' situational interest, engagement, and learning performance: The case of a physics laboratory learning environment. Journal of Science Education and Technology, 32, 1–12.

Tepe, T. (2016). New trends in educational technologies virtual reality applications. In 10th International Computer and Instructional Technologies Symposium (ICITS).

Tuna, A., & Öskan, D. (2022). A new reality in nursing students' learning: Virtual reality. The Journal of International Educational Sciences, 32(32), 320–331. https://doi.org/10.29228/inesjournal.64133

Turan, S. (2017). Akülü engelli araçları için engel algılayan çizgiler arasında giden robot tasarımı ve gerçeklenmesi. Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6(Özel Sayı ISMSIT2017), 21–29. https://izlik.org/JA24CA99PK

Wang, Y., & Liu, Y. (2024). Construction of a virtual simulation practical teaching system for intelligent manufacturing under the background of new engineering. Computer Applications in Engineering Education, e22768.

Wu, W.-L., Hsu, Y., Yang, Q.-F., Chen, J.-J., & Jong, M. S.-Y. (2023). Effects of the self-regulated strategy within the context of spherical video-based virtual reality on students' learning performances in an art history class. Interactive Learning Environments, 31(4), 2244–2267. https://doi.org/10.1080/10494820.2021.1878231

Wu, W.-Y., Guo, J.-Y., Li, Y.-J., & Sun, Y.-L. (2023). Research on the design of virtual reality online education information presentation based on multi-sensory cognition. Inventions, 8(2), 63. https://doi.org/10.3390/inventions8020063

Yolcu, V., & Demirer, V. (2017). A review on the studies about the use of robotic technologies in education. SDU International Journal of Educational Studies, 4(2), 127–139.

Yu, Z., & Xu, W. (2022). A meta-analysis and systematic review of the effect of virtual reality technology on users' learning outcomes. Computer Applications in Engineering Education, 30(5), 1470–1484. https://doi.org/10.1002/cae.22532

# AUTOMATIC CONTRAST ENHANCEMENT OF UAV IMAGERY USING DEEP LEARNING FOR IMPROVED 3D TERRAIN RECONSTRUCTION: CONTEXT OF MINING SITES APPLICATIONS

***Khaoula Abkari [1*], Samira Ben Ahmed [2] , Sara Ait-Lamallam [3] ,Souhail Kellouch [4]***

*1 Department of Cartography and Photogrammetry, College of Geomatics and Surveying Engineering, Agronomic and Veterinary Institute Hassan II, Rabat 10101, Morocco;*

*2 Department of Cartography and Photogrammetry, College of Geomatics and Surveying Engineering, Agronomic and Veterinary institute Hassan II, Rabat 10101, Morocco;*

*3 Department of Cartography and Photogrammetry, College of Geomatics and Surveying Engineering, Agronomic and Veterinary Institute Hassan II, Rabat 10101, Morocco;*

*4 AXIGEO, Surveying Company, Marrakech 40160,*

## Abstract

Reliable mining site applications based on UAV photogrammetry necessitate highly accurate 3D terrain reconstructions. However, aerial imagery acquired in these environments is notably characterized by low contrast, which constrains UAV photogrammetry through weak feature matching, leading to outputs that often fail to meet stringent accuracy requirements. This study develops a deep learning-based approach to automate contrast enhancement specifically for UAV imagery in the context of mining site applications. The proposed two-phase pipeline first employs a machine learning model (XGBoost) to predict optimal CLAHE parameters and generate a reference training dataset. A deep neural network, with a pretrained ResNet-18 encoder and a custom decoder, is then trained to perform local contrast enhancement while preserving geometric structures. Data from a reference mining site is used for absolute validation of the results. The proposed approach significantly reduces the vertical RMSE of checkpoints by 75% compared to raw images and by 67.6% compared to CLAHE. Based on point cloud, DTM, and orthophoto analyses, the results indicate that the proposed approach can improve the reliability of UAV-derived photogrammetric products under the tested mining-site conditions. Since volumetric computation and terrain monitoring accuracy are strongly dependent on DTM quality, the achieved DTM improvement suggests potential benefits for workflows targeting volume-error ranges of approximately ±3–5%, as reported in the literature. The checkpoint and orthophoto results also indicate improved compliance with high ASPRS accuracy classes.

**Keywords:** UAV photogrammetry; Low contrast; Machine Learning; Deep Learning; 3D

## 1. INTRODUCTION

UAV photogrammetry is increasingly used for mining sites applications such as monitoring, volumetric calculation, and deformation analysis. However, one of the main challenges remains the 3D reconstruction of low-contrast environments (Westoby et al., 2012; Gaou et al., 2025) to achieve these accuracy demanding works.

According to the American Society for Photogrammetry and Remote Sensing (ASPRS) standards (Photogrammetric Engineering & Remote Sensing, 2015), geospatial products are expected to comply with defined vertical and horizontal accuracy classes depending on the project requirements. In these standards, “Highest Accuracy Works” demand first classes accuracy for elevation data and orthophotos.
In mining-site monitoring and volumetric calculation, the accuracy of the Digital Terrain Model (DTM) derived from UAV imagery is a decisive factor influencing the reliability of calculated volumes. Several studies demonstrate that UAV photogrammetry can achieve centimeter-level accuracy under favorable conditions when supported by adequate ground control distributions (Matsimbe et al., 2022; Lu et al., 2023).

For most operational contexts, a vertical Root Mean Square Error (RMSE) of 5-10 cm or better is generally considered acceptable for volumetric computations and terrain monitoring, providing volume errors within approximately ±3-5 % (Tao et al., 2025). However, to meet industry standards for high-accuracy stockpile or pit monitoring, where errors below ±3 % are desired, a minimum vertical RMSE of about 0.02 m (2 cm) is recommended (Tao et al., 2025). Studies using Real Time Kinematic (RTK)/ Post-Processing (PPK) UAV and well-distributed ground control points have shown that such precision enables volumetric accuracy in the range of 1-3 % (Zhang et al., 2020).

Yet, in mining sites; that are radiometrically homogeneous or low-contrast areas; achieving these

*Corresponding author: Khaoula Abkari, abkarikhaoula@iav.ac.ma

values is difficult because feature detection and multi-view matching algorithms struggle to extract reliable correspondences (Westoby et al., 2012; Gaou et al., 2025).

### *1.1 Impact of low contrast on the photogrammetric chain*

In homogeneous environments like mining sites, low image contrast severely limits photogrammetric processing by impeding reliable feature detection and matching (Westoby et al., 2012; Chen et al., 2016). This results in fewer tie points, poor spatial distribution, and a less stable bundle adjustment (Mousavi et al., 2021). The consequences cascade through the pipeline, yielding sparse point clouds, inaccurate digital terrain models, and increased georeferencing errors (Gaou et al., 2025; Motayyeb et al., 2022). In extreme cases, such as fresh snow, reconstruction can become impossible (Gindraux et al., 2017).

### *1.2 Classical contrast enhancement methods*

Classical techniques like Histogram Equalization (HE) are often used for contrast enhancement. However, these techniques cause over-enhancement and detail loss (Pizer et al., 1986; Harichandana et al., 2020). Adaptive methods like Contrast Limited Adaptive Histogram Equalization (CLAHE) improve local contrast (Harichandana et al., 2020) but are highly sensitive to manual hyperparameter selection (clip limit, tile size) (Campos et al., 2019). Improper settings introduce artifacts or insufficient enhancement, often degrading photogrammetric results instead of improving them (Campos et al., 2019)

### *1.3 AI-Driven strategies for contrast enhancement*

Deep learning (DL) offers a data-driven alternative for image enhancement by learning radiometric transformations directly from image data. Architectures such as Zero-DCE (Guo et al., 2020) and RetinexNet (Wei et al., 2018; Hai et al., 2023) mainly address illumination correction and low-light enhancement, while MIRNet (Zamir et al., 2020; Rath et al., 2024) focuses on multi-scale image restoration and detail recovery. Although these models have shown strong performance in general enhancement tasks, their objectives are not specifically oriented toward contrast enhancement for UAV photogrammetric reconstruction. In this context, U-Net-based encoder-decoder architectures are relevant because they can be adapted to local contrast enhancement while preserving spatial and geometric structures required for 3D reconstruction.

The main characteristics of these models are summarized in Table 1. It provides a comparative overview of DL approaches for image enhancement.

**Table 1 - Deep learning models for image enhancement.**

| Model | Principle | Limitations for UAV Photogrammetry |
|---|---|---|
| Zero-DCE (Guo et al., 2020) | Uses an unsupervised Convolutional Neural Network (CNN) to perform pixel-wise light enhancement curve estimation without ground-truth reference | Only brightens images without creating new sharp gradients needed for feature matching. Not trained on UAV imagery. |
| MIRNet (Zamir et al., 2020; Rath et al., 2024) | Uses a supervised CNN with multi-scale residual learning and attention fusion for image restoration tasks | Focuses more on luminance and denoising than on local contrast. Requires large-scale training data and high GPU memory. |
| Retinex-Net (Wei et al., 2018; Hai et al., 2023) | Uses a supervised CNN based on Retinex theory to separate illumination and reflectance to enhance contrast and remove noise | Primarily designed for terrestrial scenes or facial images under low light, not for homogeneous UAV surfaces under normal lighting. |
| SICE (Cai et al., 2018) | Uses a supervised CNN to learn deep contrast enhancement from multi-exposure image sequences | Requires multi-exposure data, which is not typical for UAV mapping flights. Built on an older framework (Caffe), limiting compatibility. |
| U-Net (Ehab, Huang et al., 2024) | Uses a supervised U-shaped encoder-decoder architecture. | Originally designed for segmentation. |

Contrast enhancement in UAV imagery has long relied on traditional image processing techniques. The most common approaches include Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast-Limited Adaptive Histogram Equalization (CLAHE), Linear Contrast Stretch (Pizer et al., 1986; Harichandana et al., 2020) and Gamma Correction (Rahman et al., 2016; Xu et al., 2020). These techniques, widely integrated into image processing and photogrammetry software, aim to redistribute pixel intensities and facilitate the detection of homologous points. Classical methods, although simple and integrated, have structural limitations in low-texture scenes. Global histogram equalization enhances overall contrast but often causes over-enhancement and loss of fine details (Touhami et al., 2024). Adaptive Histogram Equalization (AHE) can introduce artificial contrast structures and amplify noise in homogeneous regions, producing visually unrealistic outputs (Gomes, 2008). Contrast Limited Adaptive Histogram Equalization (CLAHE), while mitigating some of these issues, remains highly sensitive to parameters such as tile size and clip limit, which require manual tuning; incorrect settings may lead to artifacts and noise amplification (Campos et al., 2019; Harichandana et al., 2020). When many tiles are used, CLAHE can consume most of the available

resources, making it unsuitable for integration into real-time or embedded vision systems (Kryjak et al., 2022).

With the rapid development of artificial intelligence, several deep learning-based models have been introduced to address limitations associated with classical image enhancement techniques. Among them, Zero-DCE (Guo et al., 2020) proposes an unsupervised learning framework that estimates pixel-wise curve parameters to adjust illumination without the need for reference images. Similarly, Retinex-Net (Wei et al., 2018; Hai et al., 2023) decomposes images into reflectance and illumination layers, allowing adaptive adjustment of light intensity and improved visibility in under- or overexposed regions. These approaches are mainly oriented toward illumination correction and visibility improvement, which makes their objectives different from contrast enhancement evaluated through photogrammetric reconstruction quality. Other architectures, such as MIRNet (Zamir et al., 2020; Rath et al., 2024) and U-Net (Majidizadeh, Hasani et al., 2023; Ehab, Huang et al., 2024), employ encoder-decoder structures with skip connections to enhance local details, edges, and textures while maintaining overall luminance balance. The SICE framework (Cai et al., 2018) explores multi-exposure fusion through convolutional neural networks to reconstruct well-exposed images from a set of low-quality inputs. More recent studies have extended these architectures to hybrid or attention-based models, enabling better feature extraction and global-local adaptation of brightness and contrast. However, these models are generally evaluated using image-quality or perceptual criteria, whereas UAV photogrammetric applications require assessing whether image enhancement improves feature matching, bundle adjustment stability, and the accuracy of derived products.
Therefore, the relevance of deep learning-based enhancement for UAV photogrammetry cannot be assessed only through visual image quality. It also requires photogrammetric indicators such as tie point density, reprojection error, camera calibration stability, checkpoint accuracy, point cloud consistency, DTM quality, and orthophoto accuracy.

This study follows this perspective by developing a deep learning-based framework for automatic contrast enhancement of UAV images in homogeneous mining-site environments, where weak radiometric variations can limit feature detection, image matching, and 3D reconstruction reliability. The proposed approach integrates a machine learning stage to predict optimal CLAHE parameters and generate a task-specific reference training dataset. It also integrates a convolutional neural network that performs adaptive local enhancement while preserving geometric structures relevant to photogrammetric reconstruction. Its contribution lies in combining ML-based parameter optimization with DL-based local contrast enhancement and in evaluating the resulting images through photogrammetric indicators, including tie point density, reprojection error, camera calibration stability, checkpoint accuracy, point cloud consistency, DTM quality, and orthophoto accuracy.

## 2. MATERIALS AND METHODS

### *2.1 General methodology*

The proposed methodology, illustrated in Figure 1 diagram, provides a detailed overview of the four main stages of the adopted general methodology. First, data preprocessing and preparation, based on raw aerial images, is achieved. Then, data portioning is done to generate training, test and validating data. After, model training using a U-Net customized architecture is elaborated. Finally, results evaluation is achieved based on comparative analyses. Section 2.4 gives detailed methodology steps.

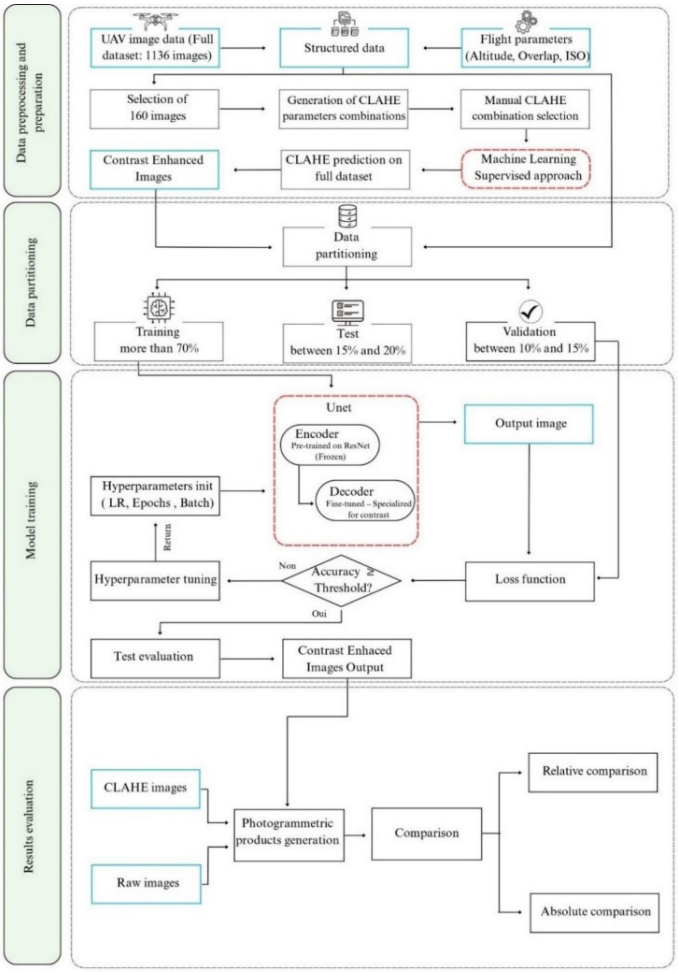
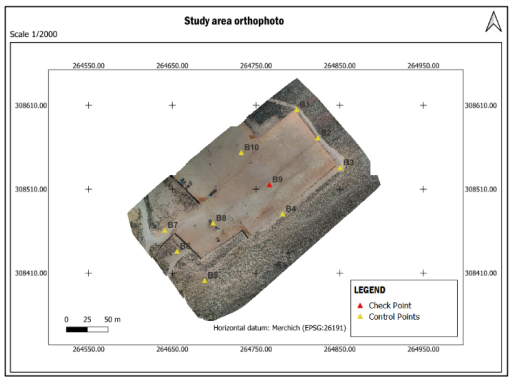



**Figure 1**: *Methodology diagram*

### *2.2 Study and reference areas and datasets*

The study area is a mining site in the province of Berrechid in Morocco. It's a quarry of 40000 m²

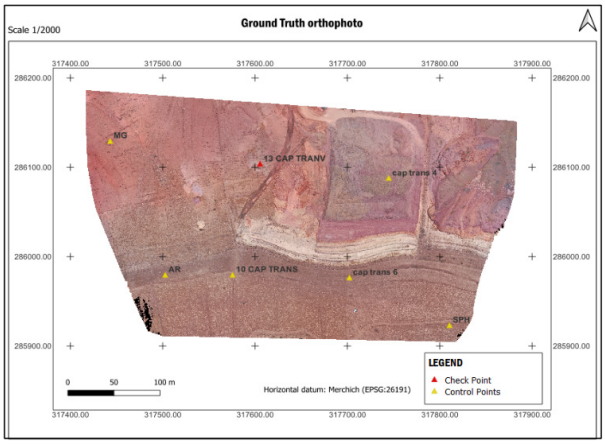
surface. The site produces construction aggregates and is among the most important mining sites in the country. The terrain is characterized by predominantly calcareous mineral composition. As described in Figure 2. The soil exhibits a homogeneous
texture and low contrast in aerial images. The aerial images acquired in the study area served to experiment the methodology. The reference area is another mining site in Morocco. It is located in the province of Berrechid too, about 25 km from the city center. Its location is presented in Figure 3. Its surface is 110000 m². This site is selected because it presents similar characteristics as the study site. It serves to validate the results of the study by comparative analyses.




**(a)**

**(b)**

**Figure 2**: *Geographic overview of the study: (a) Geographic location and extent of the study area; (b) zoom on the soil illustrating the surface homogeneity.*




**(a)**

**(b)**

**Figure 3:** *Geographic overview of the reference area: (a) Geographic location and extent of the reference area; (b) zoom on the soil illustrating the surface homogeneity.*

### *2.3 Reference datasets*

In this paper, Reference Data 1 designates the dataset derived from the application of the CLAHE algorithm. This dataset includes UAV images acquired on the study site. Seven flight missions are conducted under varying conditions of flight altitude, overlap, ISO sensitivity, and spatial resolution (GSD). In addition, 9 Ground Control Points (GCPs) are used for georeferencing, among which point "B9" is specifically designated as a checkpoint. This type of reference data is used for relative validation, as it enables a fair comparison between the original UAV images, the CLAHE-enhanced ones, and those enhanced by DL.On the other hand, Reference Data 2 designates the dataset acquired on the reference site. This dataset includes UAV images collected during a mission at a flight altitude of 90 m, with an image overlap of 80%, an ISO value set to Automatic (100), and a spatial resolution (GSD) of 3.15 cm/pixel. In addition, 7 Ground Control Points (GCPs) are used for georeferencing, among which point "13 CAP TRANV" is specifically designated as a checkpoint. This second type of reference data is used for absolute validation, as it allows for the evaluation of georeferencing accuracy and spatial reliability against a physically verified site.

The characteristics of the two reference datasets are summarized in Table 2.

| Reference dataset | Validation type | Data basis | Associated site |
|---|---|---|---|
| Reference Data 1 | Relative validation | UAV images from 7 missions at the study area; 9 GCPs (checkpoint: B9) | Study area |
| Reference Data 2 | Absolute validation | UAV images at the reference site; 7 GCPs (checkpoint: 13 CAP TRANV) | Reference site |

### *2.4 Detailed Methodology*

#### *2.4.1 Data processing and preparation*

Reference images are generated by applying CLAHE to raw UAV inputs. The hyperparameters, namely clip limit and tile number, are optimally selected through the XGBoost supervised classification model (Chen et al., 2016), as shown in Figure 4. This figure illustrates the workflow of the ground truth generation process. In this study, the generated ground truth refers to a task-specific reference derived from ML-optimized CLAHE outputs, rather than an independent radiometric ground truth. This enabled the automatic creation of contrast-enhanced image pairs for neural network training.

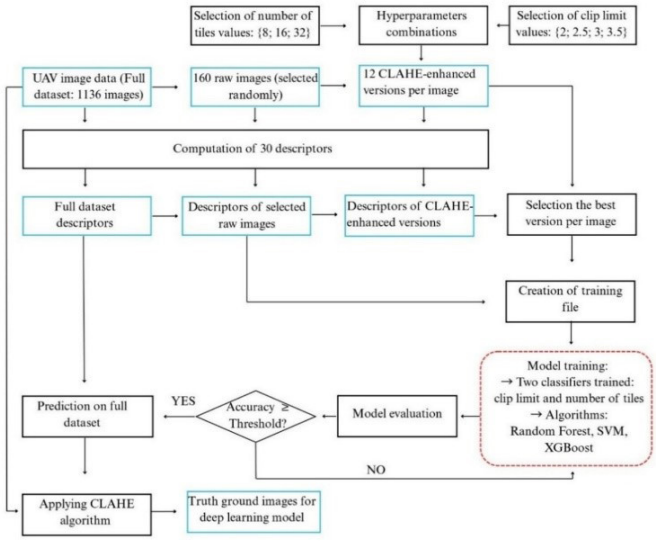



**Figure 4**: *Workflow of the Ground truth generation process*

#### *2.4.2 Data partitioning*

Dataset, comprising 1136 UAV images with varying flight altitudes, overlaps, and ISO settings, is partitioned into training, validation, and test sets. The same splits are applied to the corresponding machine learning–enhanced ground truth images.
Data preprocessing included augmentation. Random flipping and rotation are included to increase training diversity. Conversion of each image into a PyTorch tensor (Paszke et al., 2019) for neural network input followed.

#### *2.4.3 Model training*

A customized U-Net is used for UAV image contrast enhancement. A frozen pre-trained encoder for feature extraction and a custom decoder to generate contrast-enhanced outputs are integrated.
The overall neural network architecture used for contrast enhancement is illustrated in Figure 5

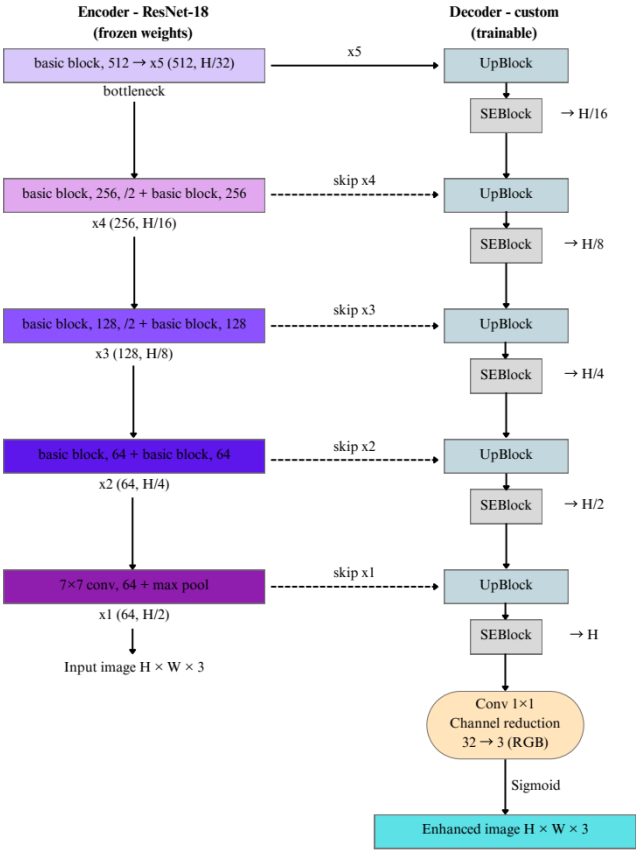



**Figure 5**: *Overall architecture of the proposed contrast-aware U-Net*

Feature extraction mechanism relied on pretrained ResNet-18 and ResNet-50 (De Camargo et al., 2021) backbones. It captures multi-scale intensity variations as illustrated in Figure 6. Both encoders, with frozen weights, provided hierarchical feature maps via skip connections to the decoder for contrast enhancement

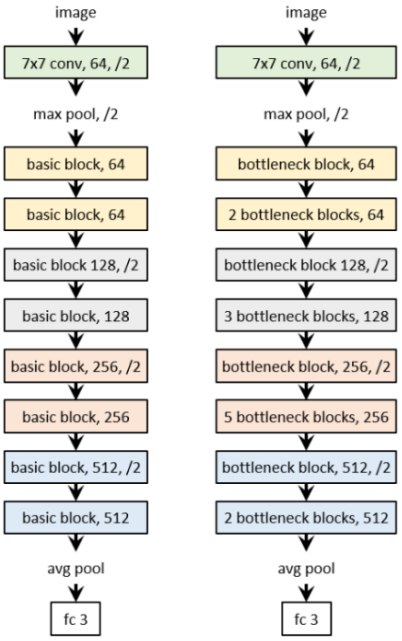



**Figure 6**: *ResNet-18 and ResNet-50 architectures. (Left): ResNet-18. (Right): ResNet-50 (Choi et al., 2020).*

For image reconstruction, the decoder reconstructs contrast-enhanced images from hierarchical encoder features through four up-sampling blocks. A combination of transposed convolutions, skip connection fusion, convolutional processing with batch normalization and Exponential Linear Unit (ELU) (Barron, 2017), and channel attention via SEBlocks is realized. Figure 7 illustrates the decoder architecture of the contrast-aware U-Net. The final 1×1 convolution and sigmoid activation produce the normalized output..

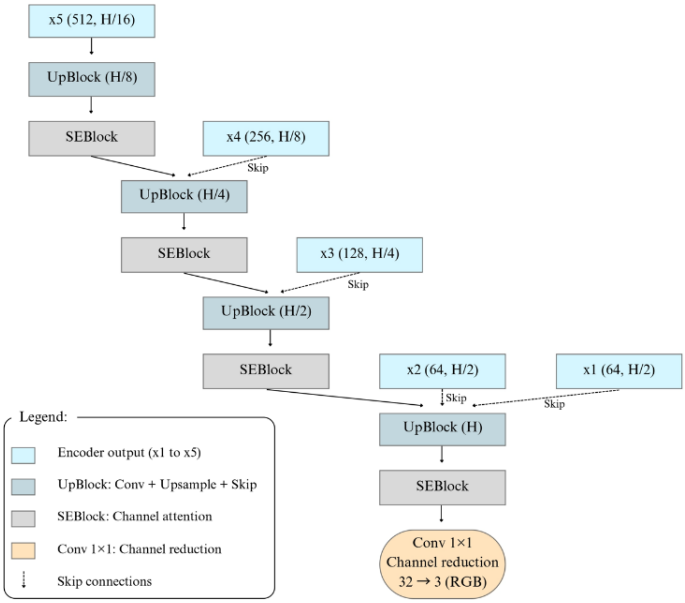



**Figure 7**: *Decoder architecture of the contrast-aware U-Net.*

For training functions, two loss functions are tested: a composite L1- Gradient loss to balance intensity accuracy and detail preservation, and Structural Similarity Index Measure (SSIM) (Nilsson and Akenine-Möller, 2020) loss to capture perceptual similarity. Optimization is performed using the Adam algorithm (Kingma and Ba, 2017) for efficient convergence.

The training experiments were conducted on a high-performance workstation equipped with an AMD Ryzen 9 7950X3D 16-Core Processor, 128 GB DDR5 RAM, an NVIDIA GeForce RTX 4060 Ti GPU with 8 GB memory, and a 4 TB NVMe SSD. The models were implemented in Python using PyTorch within an Anaconda environment. Training duration varied depending on the encoder architecture and hyperparameter configuration. The selected ResNet-18 configuration, trained for 25 epochs with a batch size of 2 and a learning rate of $10^{-3}$, required approximately 3327 seconds, while the ResNet-50 configuration under comparable conditions required approximately 23266 seconds. This confirmed that ResNet-18 provided the best compromise between reconstruction quality and computational efficiency.

### *2.4.4 Model evaluation*

The evaluation of the model's contrast enhancement performance is based on three standard image quality metrics: Mean Squared Error (MSE) (Marmolin, 1986), Peak Signal-to-Noise Ratio (PSNR) (Tanchenko, 2014). and Structural Similarity Index Measure (SSIM) (Nilsson and Akenine-Möller, 2020). Table 3 presents the role of each metric in the evaluation of the results.

| Metric | Evaluates |
|---|---|
| MSE | Pixel-wise distortion |
| PSNR | Overall distortion in decibels (dB) |
| SSIM | Perceptual quality (structure, luminance and contrast) |

**Table 3**: *Role of Evaluation Metrics in Result Assessment*

### *2.5 Results evaluation*

### *2.5.1 Relative evaluation*

The impact of contrast enhancement is assessed by comparing deep learning-enhanced, CLAHE, and raw images using photogrammetric metrics.
These metrics are:
• Tie points number;
• Reprojection error;
• Camera calibration parameters
• GCP RMSE.
They are used to evaluate improvements in aerotriangulation accuracy.

### *2.5.2 Absolute evaluation*

Absolute evaluation is conducted against ground-truth data by comparing RMSE of check points, 3D points clouds, Digital Terrain Models (DTM) and orthophotos. The accuracy is assessed through RMSE of selected control points, to evaluate the accuracy of CLAHE and deep learning-enhanced images relative to reference photogrammetric products.

### *2.6 Materials*

To achieve aerial missions, the UAV used is the DJI Mavic 2 Pro (DJI Mavic 2, 2025). It is a multirotor UAV equipped with a Hasselblad camera featuring a 1-inch 20 MP CMOS sensor (Support for Mavic 2, 2025). Table 4 illustrates the technical specifications of the sensor.

| Camera sensor | Technical specifications |
|---|---|
| Type | 1-inch CMOS, 20 MP model FC6310S |
| Focal Length | 35 mm equivalent format: 28 mm |
| FOV | approx. 77° |

| | |
|---|---|
| Aperture | f/2.8 – f/11 |
| Focus | 1 m to ∞ |
| ISO Range | 100–3200 (auto) / 100–12800 (manual) |
| Shutter Speed | Electronic shutter: 8 s to 1/8000 s |
| Image Size | 5472 × 3648 |
| Image Format | JPEG/DNG (RAW) |

**Table 4**: Overview of Camera Sensor Specifications

To achieve the methodology steps, several software are used. ArcGIS Pro (ArcGIS Pro, 2025) (version 3.4) is used for the classification of dense point clouds to separate ground from non-ground objects, in order to generate a Digital Terrain Model (DTM). QGIS (QGIS, 2025) (version 3.28) is used for the evaluation of Digital Terrain Models (DTMs) by computing raster differences between the reference DTM and those generated from images enhanced using CLAHE and DL. Pix4DMapper (PIX4Dmapper, 2025) (version 4.5.6) is used for UAV image processing, including the generation of point clouds, Digital Elevation Models (DEMs), and orthophotos. Cloud Compare (CloudCompare, 2025) (version 2.13) is used for the evaluation of 3D point clouds using the Cloud-to-Cloud distance (C2C) plugin. Anaconda (Anaconda, 2025) version 2024.06 and Visual Studio Code (Visual Studio Code, 2025) (version 1.90) are used for executing Python scripts related to contrast enhancement, as well as the training of DL and machine learning models. Table 5 provides a description of the software and libraries used in this study.

| Software/ Platform | Library |
|---|---|
| QGIS | PyTorch |
| Anaconda | OpenCV (cv2) |
| Visual Studio Code | NumPy |
| Pix4DMapper | Pillow |
| Cloud Compare | Pandas |
| ArcGIS Pro | Scikit-learn |
| ArcGIS Pro | Joblib |
| ArcGIS Pro | Scipy |
| ArcGIS Pro | Matplotlib |

**Table 5**: Description of the software and library used

# 3. RESULTS

## *3.1 Machine Learning prediction of CLAHE parameters*

The automatic selection of optimal CLAHE parameters is the foundational step for generating the training dataset. The XGBoost model demonstrates superior performance in predicting the optimal hyperparameters compared to other algorithms, as illustrated in Figure 8. The figure compares the accuracy, precision, recall, and F1-score for both clip limit prediction and number of tiles prediction for the three machine learning algorithms used.

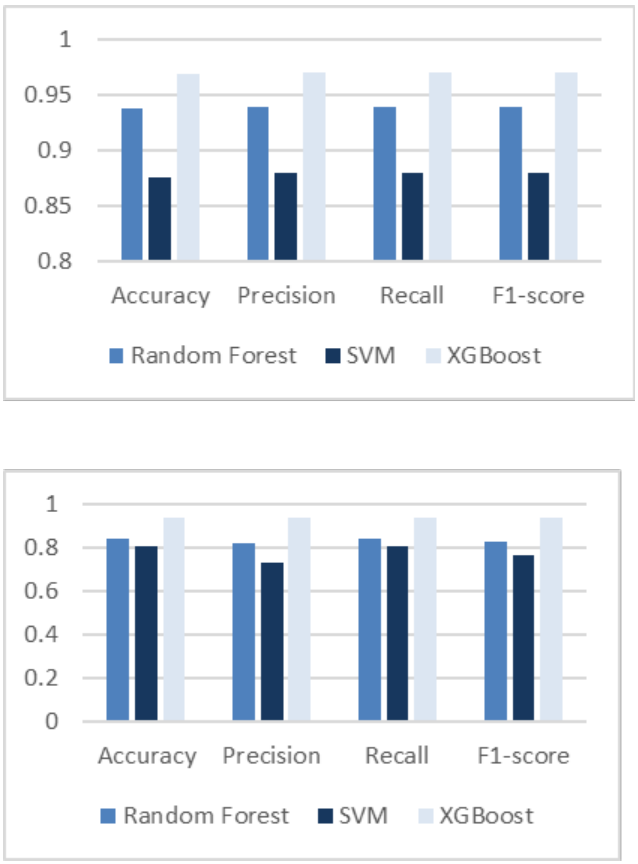



**Figure 8**: *Performance comparison of machine learning models for predicting CLAHE hyperparameters: (a) accuracy, precision, recall, and F1-score for clip limit prediction; (b) accuracy, precision, recall, and F1-score for the number of tiles prediction.*

## *3.2 Model evaluation*

### *3.2.1 Training and validation evaluation*

The UNet model is customized through a comprehensive experimental framework. Encoder architectures, loss functions, and hyperparameter configurations are evaluated. The complete experimental matrix consisted of 21 unique configurations, with key results presented in Table 6.

| Encoder | Loss function | Epochs | Batch size | Learning rate | Train loss |
|---|---|---|---|---|---|
| ResNet-18 | Composite | 25 | 2 | $10^{-3}$ | 0,0517 |
| ResNet-50 | Composite | 25 | 2 | $10^{-3}$ | 0,0492 |
| ResNet-18 | SSIM | 25 | 4 | $10^{-4}$ | 0,0517 |
| ResNet-18 | Composite | 50 | 4 | $10^{-3}$ | 0,0441 |

**Table 6**: *Performance comparison of representative UNet configurations on validation metrics.*

| Encoder | Validation loss | PSNR (dB) | SSIM | MSE | Duration (s) |
|---|---|---|---|---|---|
| ResNet-18 | 0,0459 | 26,32 | 0,9712 | 0,0028 | 3327 |
| ResNet-50 | 0,0482 | 26,49 | 0,9697 | 0,0028 | 23266 |
| ResNet-18 | 0,0483 | 20,63 | 0,9516 | 0,0105 | 3564 |
| ResNet-18 | 0,0430 | 27,56 | 0,9807 | 0,0024 | 10240 |

**Table 6** *(continued) :Performance comparison of representative UNet configurations on validation metric*

Experimental results indicate four primary findings:

• The configuration with ResNet-18 encoder, composite loss function, 25 epochs, batch size 2, and learning rate $10^{-3}$ achieved optimal balance between performance and computational efficiency

• ResNet-50 provided comparable performance (MSE: 0,0028) under identical conditions but required substantially longer training time (23266 seconds vs 3327 seconds)

• The composite loss function consistently outperformed SSIM loss across all configurations, with significantly lower MSE values (0,0028 vs 0,0105)

• Training and validation loss metrics show consistent correlation with reconstruction quality metrics (PSNR, SSIM and MSE).

Through this experimentation, an optimal configuration using ResNet-18 is adopted. It is illustrated in Figure 9.

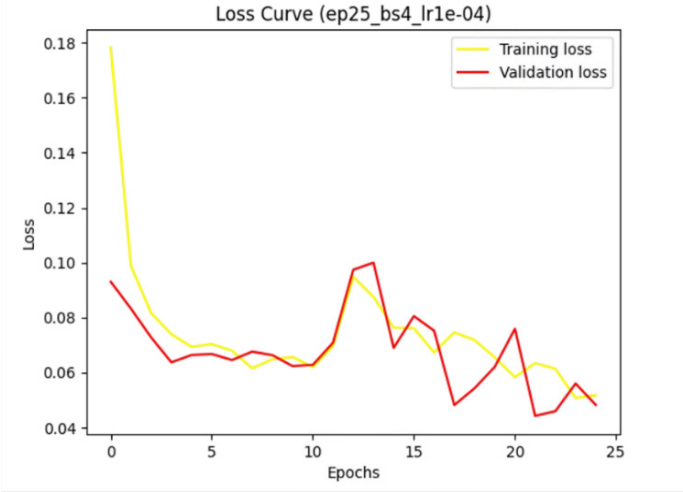



**Figure 9:** *Training and validation loss curves for UNet with ResNet-18 using SSIM loss (batch = 4, lr = $10^{-4}$, 25 epochs)*

### *3.2.2 Test evaluation*

The generalization capability of the optimized model is evaluated on unseen test data. The results are presented in Table 7. The SSIM value of 0.9581 indicates high structural similarity between processed and reference images. The PSNR value of 25.82 dB corresponds to favorable reconstruction quality, while the MSE of 0.0027 represents low pixel-wise error. These metrics demonstrate the model's ability to maintain image quality characteristics when processing previously unseen UAV imagery.

| Metric | SSIM | PSNR | MSE |
|---|---|---|---|
| Value | 0,9581 | 25,82 | 0,0027 |

**Table 7**: *Test evaluation metrics of the final model*

### *3.3 Relative result evaluation*

### *3.3.1 Visual comparison*

Visual assessment demonstrates significant improvement in image quality through contrast enhancement methods. The raw image, shown in Figure 10, exhibits low contrast with limited visible detail. CLAHE enhanced image, presented in Figure 11, illustrates the algorithm's ability to improve contrast and fine details appearing much more distinct. DL enhancement, illustrated by Figure 12, provides optimal contrast improvement, the difference between neighboring pixels is more pronounced.

**Figure 10**: *Raw image*

**Figure 11**: *CLAHE contrast enhanced image*

**Figure 12**: *Deep learning contrast enhanced image*

### *3.3.2 Tie points density and reprojection error*

CLAHE enhancement increases the number of tie points (+31 to 33 %) but slightly raises the reprojection error (+6 %). In contrast, DL yields a much larger gain in tie points (+56 to -78 %) while simultaneously reducing reprojection error (-12 to -19 %). Table 8 highlights these results.

| Dataset | Photogrammetric Metric | Raw images | Contrast-enhanced images (CLAHE) | Contrast-enhanced images (DL) |
|---|---|---|---|---|
| Dataset 1 | Number of Tie Points | 2642289 | 3518573 | 4123400 |
| | Reprojection error (px) | 0.197 | 0.210 | 0.169 |
| Dataset 2 | Number of Tie Points | 1184307 | 1552776 | 2109142 |
| | Reprojection error (px) | 0.185 | 0.196 | 0.162 |

**Table 8:** *Tie points and reprojection error comparison.*

### *3.3.3 Camera Calibration Parameters*

Contrast enhancement improves camera calibration, with CLAHE slightly reducing focal length error (Dataset 1: 0.006 mm to 0.005 mm; Dataset 2: 0.003 mm) and DL further reducing it (Dataset 1: 0.005 mm; Dataset 2: 0.002 mm). DL also decreases camera position errors and orientation errors, providing more accurate and reliable calibration, particularly in the lower-contrast Dataset 2. Figure 13 and Figure 14 illustrate these outcomes.

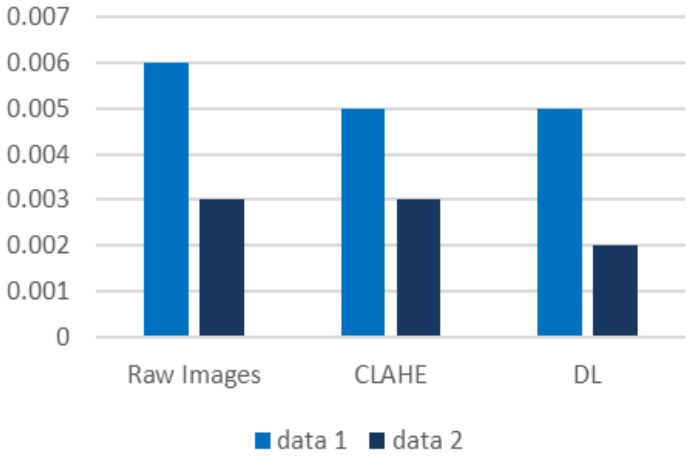



**Figure 13:** *Focal length error (mm).*

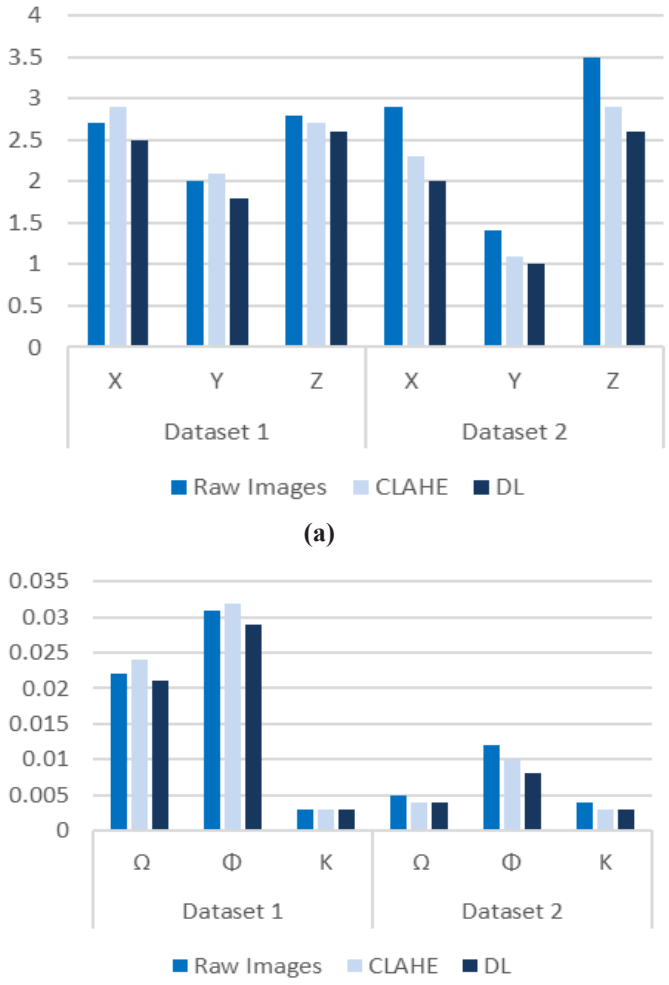



**Figure 14:** *Comparison of errors for the two datasets: (a) Camera position error (cm), (b) Camera orientation error (°)*

### *3.3.4 Ground Control Points RMSE*

The internal accuracy of the bundle adjustment, quantified by GCP RMSE values shows clear differences across processing methods. Raw imagery of Dataset 2 exhibits excessive vertical error (RMSEz = 18 cm), indicating potential blunders according to ASPRS guidelines. In contrast, DL enhanced data reduces total RMSE to 3.0 cm, demonstrating optimal internal consistency and providing a reliable basis for subsequent absolute accuracy validation with independent check points. Table 9 gives detailed results on GCP RMSE.

| Dataset | UAV images | RMSEx (cm) | RMSEy (cm) | RMSEz (cm) | Total RMSE (cm) |
|---|---|---|---|---|---|
| Dataset 1 | Raw images | 1.80 | 0.70 | 1.50 | 2.40 |
| | Contrast-Enhanced images (CLAHE) | 1.60 | 0.60 | 1.40 | 2.20 |
| | Contrast-Enhanced images (DL) | 0.90 | 0.40 | 0.60 | 1.10 |
| Dataset 2 | Raw images | 2.90 | 2.30 | 17.60 | 18.00 |
| | Contrast-Enhanced images (CLAHE) | 1.90 | 2.30 | 8.40 | 8.90 |
| | Contrast-Enhanced images (DL) | 0.90 | 1.00 | 2.70 | 3.00 |

**Table 9**: *Results on the RMSE of the ground control points (GCP).*

### *3.4 Absolute result evaluation*

#### *3.4.1 Check point RMSE*

Results on horizontal and vertical accuracy assessments based on independent GNSS checkpoints are realized. These accuracies are also compared to the ASPRS Positional Accuracy Standards. The results allow a direct comparison between raw UAV imagery, CLAHE enhancement, and the proposed DL-based method. Table 10 and Table 11 summarize these outcomes.

| Dataset | Product | RMSEx | RMSEy | RMSEr | ASPRS Horizontal Class |
|---|---|---|---|---|---|
| 1 | Raw Images | 3.70 | 4.70 | 8.50 | 5.0 cm (≤ 10.0) |
| | CLAHE | 1.10 | 1.90 | 5.50 | 5.0 cm (≤ 10.0) |
| | DL | 0.90 | 0.40 | 1.10 | 0.63 cm (≤ 1.30) |
| 2 | Raw Images | 6.80 | 3.60 | 12.30 | 7.5 cm (≤ 15.0) |
| | CLAHE | 2.60 | 2.30 | 8.20 | 5.0 cm (≤ 10.0) |
| | DL | 2.00 | 1.40 | 3.40 | 2.5 cm (≤ 5.0) |

**Table 10**: *Horizontal accuracy evaluation from independent GNSS checkpoints (values in cm).*

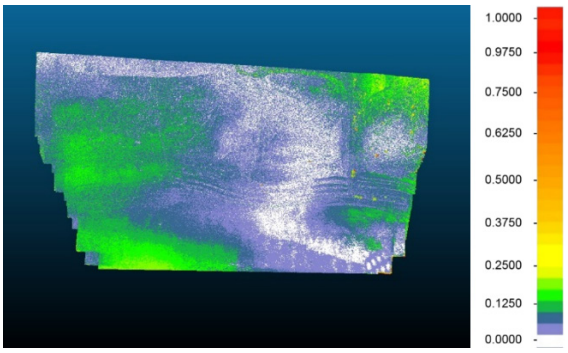
| Dataset | Product | RMSEz | NVA (95%) | ASPRS Vertical Class |
|---|---|---|---|---|
| 1 | Raw Images | 6.10 | 11.96 | 10 cm (5 < RMSEz ≤ 10) |
| | CLAHE | 5.10 | 10.00 | 10 cm (5 < RMSEz ≤ 10) |
| | DL | 0.50 | 0.98 | ≤ 5 cm |
| 2 | Raw Images | 9.60 | 18.82 | 10 cm (RMSEz ≤ 10) |
| | CLAHE | 7.40 | 14.50 | 10 cm (RMSEz ≤ 10) |
| | DL | 2.40 | 4.70 | 5 cm (RMSEz ≤ 5) |

**Table 11**: *Vertical accuracy evaluation from independent GNSS checkpoints (values in cm).*

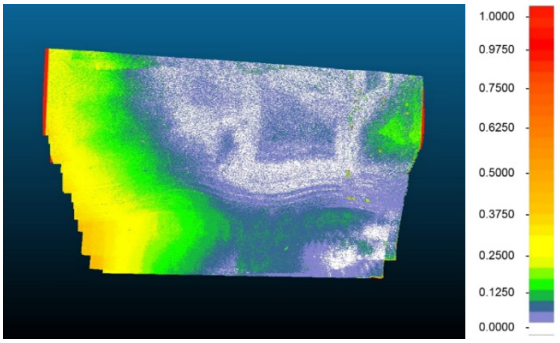
The results indicate that the deep learning (DL) approach improves positional accuracy and yields a substantial reduction in vertical RMSE across both reference datasets. According to the ASPRS Positional Accuracy Standards (Photogrammetric Engineering & Remote Sensing, 2015), the DL-enhanced products reached higher accuracy classes than the raw and CLAHE-based products under the tested conditions. Specifically, for Dataset 2, the proposed method reduces the vertical RMSE by 75% relative to the raw images and by 67.6% compared to CLAHE. These outcomes support the relevance of the proposed approach for improving UAV-derived 3D products in low-contrast mining-site environments.

#### *3.4.2 Point cloud accuracy*

Cloud-to-cloud (C2C) distance analysis is performed to evaluate point cloud accuracy. Point clouds generated from DL enhanced images are compared against those generated from CLAHE-enhanced images as well as raw images from the reference Data 2. This comparison allows quantifying the geometric deviations introduced by each image enhancement method and assessing the relative performance of the DL approach. As shown in Figure 15, the deep learning-enhanced point cloud demonstrates superior accuracy with smaller, more homogeneous deviations compared to the CLAHE-enhanced reconstruction, which exhibits larger, irregular errors.




(a)



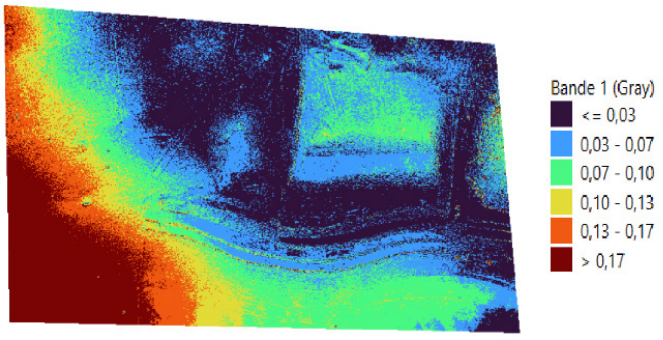

(b)

**Figure 15:** *Visual Comparison of C2C deviations between reconstructed and reference data 2 point clouds: (a) deep learning; (b) CLAHE; (values in meter)*

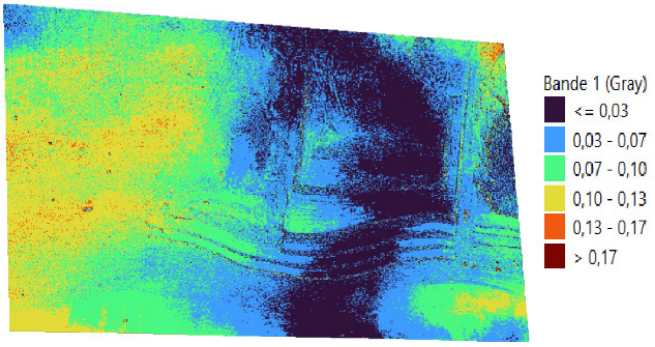
Quantitative analysis shows the CLAHE method allows 9.78 cm mean distance, with 8.48 cm standard deviation. The DL approach achieved a mean distance of 7.67 cm with a standard deviation of 3.94 cm, representing a significant improvement over CLAHE method. This 53.5% reduction in standard deviation indicates substantially higher accuracy.

### *3.4.3 Digital terrain model comparison*

The digital terrain models (DTMs) generated from each image enhancement method are evaluated through raster difference analysis against the DTM generated from reference Data 2, with the outcomes visually illustrated in Figure 16. Quantitative assessment focused on mean elevation error and standard deviation of differences.




(a)




(b)

**Figure 16**: *Elevation comparison between the DTMs generated from: (a) CLAHE-enhanced images and the reference DTM, (b) DL-enhanced images and the reference DTM; (values in meter)*

The deep learning-enhanced DTM demonstrated superior performance with a mean error of 6 cm and standard deviation of 4 cm, compared to the CLAHE-based DTM which showed higher deviations (mean error: 8 cm, standard deviation: 7 cm). Figure 17 presents these results.

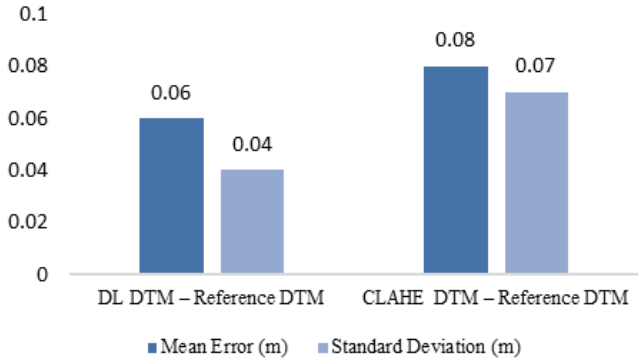



**Figure 17:** *Digital Terrain Model evaluation.*

The DL-based DTM reached the 5 cm class for Swath-to-Swath Non-Vegetated Terrain according to the ASPRS vertical accuracy standards (Photogrammetric Engineering & Remote Sensing, 2015). The DL approach also showed improved consistency, reducing the standard deviation by 43% compared to the CLAHE-based DTM. This improvement indicates more stable elevation information for mining-site applications, where relative accuracy and surface consistency are critical. Since volume and slope calculations are directly derived from DTM quality, the observed improvement in elevation consistency suggests potential benefits for volumetric computation and terrain monitoring workflows, for which volume-error ranges of approximately ±3–5% are commonly reported in operational contexts (Tao et al., 2025).

### *3.4.3 Orthophoto comparison*

The geometric accuracy of orthophotos is evaluated using five ground reference points, with comparisons

made between reconstructed orthophotos from each enhancement method and the reference orthophoto generated from reference Data 2. Figure 18 illustrates the spatial distribution of evaluation points, while Figure 19 presents the accuracy assessment results.

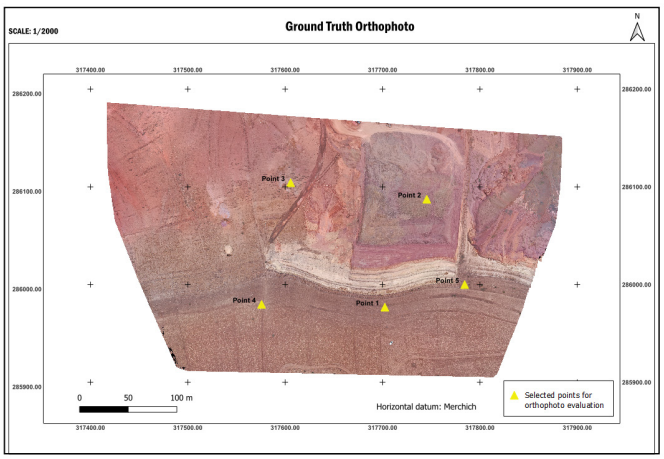
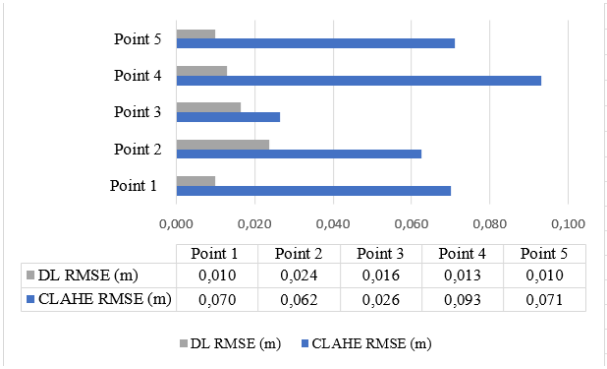
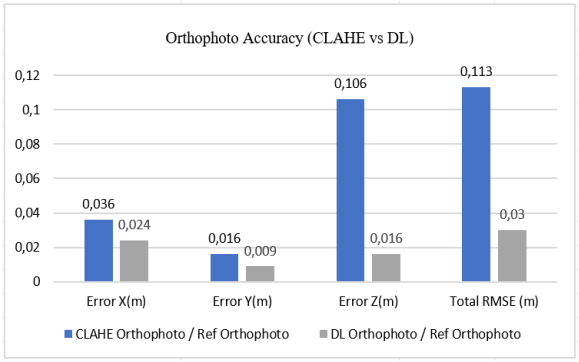



**Figure 18**: *Ground truth orthophoto-plan illustrating the spatial distribution of selected points for orthophoto evaluation.*




**(a)**




**(b)**

**Figure 19**: *Orthophoto accuracy assessment: (a) RMSE values at individual check points for both enhancement methods; (b) comparative error components along X, Y, and Z axes.*

The analysis revealed significant improvements with DL enhancement across all positional components. CLAHE-based orthophotos exhibited errors of 3.60 cm (X), 1.60 cm (Y), and 10.60 cm (Z), resulting in a total RMSE of 11.30 cm. In contrast, deep learning-enhanced orthophotos achieved substantially lower errors: 2.40 cm (X), 0.90 cm (Y), and 1.60 cm (Z), with a total RMSE of 3.00 cm.
This represents relative accuracy improvements of 33.3% in horizontal positioning (X-Y) and 84.9% in vertical accuracy (Z). The significant reduction in Z-axis error indicates enhanced elevation consistency in the orthophoto products. These outcomes indicate that the deep learning-enhanced orthophotos reached higher ASPRS accuracy classes than the CLAHE-based products under the tested conditions.

## 4. DISCUSSION

This study investigates automatic UAV imagery contrast enhancement through Machine Learning and Deep Learning approaches, aiming to improve photogrammetric reconstruction quality in the context of mining site applications. A first stage focused on predicting CLAHE hyperparameters from raw image descriptors, where XGBoost achieved the highest accuracies (0.969 for clip limit and 0.938 for tile number), confirming its robustness for automated parameter selection. A second stage explored adaptive DL-based enhancement, testing multiple encoder-decoder configurations and loss functions. Optimal models produced enhanced images without artifacts and yielded clear photogrammetric benefits. Compared to raw and CLAHE-enhanced imagery, DL-enhanced images significantly increased tie point density, reduced reprojection errors, and improved camera calibration stability. Accuracy assessments confirmed substantial reductions in planimetric and altimetric errors. Orthophotos, DTMs, and 3D point clouds generated from DL-enhanced images exhibited lower positional deviations and error dispersion than those from CLAHE, highlighting DL's superior ability to preserve geometric fidelity in homogeneous, low-contrast environments. Evaluations based on the ASPRS Positional Accuracy Standards (Photogrammetric Engineering & Remote Sensing, 2015) further support these findings. Across both datasets, DL-enhanced products reached higher ASPRS accuracy classes than raw and CLAHE-enhanced imagery. GCP-based assessments revealed high vertical errors in raw Dataset 2 (RMSEz ≈ 18 cm), whereas DL-enhanced imagery reduced this value to approximately 3 cm. Checkpoint analysis confirmed the same trend, with DL-enhanced orthophotos achieving total RMSE values around 3 cm compared to 11 cm for CLAHE, corresponding to a 33% improvement in horizontal positioning and nearly 85% in vertical accuracy. DL-based DTMs also showed lower mean vertical error (6 cm) and reduced

dispersion compared to CLAHE, which is particularly relevant for elevation-sensitive mining applications such as volume estimation and terrain monitoring, where volume-error ranges of approximately ±3–5% are commonly reported in operational contexts.

It is worth noting that the proposed approach differs fundamentally from general-purpose image enhancement models such as Zero-DCE or MIRNet, which primarily target luminance adjustment in natural or low-light scenes. The present method is specifically designed and optimized for UAV aerial imagery in low-contrast, radiometrically homogeneous environments, where the photogrammetric integrity of the output rather than perceptual quality is the primary objective. This domain-specific focus explains the architectural and training choices made, and accounts for the performance observed between the proposed approach and classical enhancement techniques in this particular context. While CLAHE improved contrast, its sensitivity to hyperparameter selection sometimes caused artifacts and spatial inaccuracies. DL-based enhancement provided more adaptive contrast adjustment than CLAHE under the tested conditions, with improved geometric accuracy and stability, particularly in the altimetric components relevant to terrain monitoring and volume-related analyses. The training strategy relies on CLAHE-enhanced images as reference data, which links the learning process to the quality of this intermediate enhancement step. However, the photogrammetric validation conducted on a separate reference site shows that the learned enhancement produces measurable geometric improvements compared to CLAHE alone. Overall, DL-based adaptive contrast enhancement showed stronger performance in this experimental context, while ML-based CLAHE parameter prediction remains a valuable intermediate approach when computational resources are constrained.

## 5. CONCLUSIONS

This study contributes to photogrammetric preprocessing by proposing an automated UAV image contrast enhancement workflow for homogeneous low-contrast areas.The main objective is to improve the quality of 3D terrain reconstruction, in the context of mining sites applications, and the use of DL proved highly effective. A fine-tuned U-Net model substantially improved reconstruction results by increasing tie point density, reducing reprojection error, enhancing camera calibration, and lowering planimetric and altimetric errors. As a result, the derived photogrammetric products, including dense point clouds, digital terrain models (DTMs), and orthophotos, demonstrated improved accuracy and reliability.

Under the tested conditions, DL-based enhancement outperformed CLAHE across the evaluated photogrammetric indicators, while offering automatic and adaptive contrast enhancement without manual hyperparameter tuning. Compared with CLAHE, whose performance remains sensitive to parameter configuration and may lead to noise amplification or artifacts, the DL approach better preserved structural details and produced more accurate photogrammetric outputs in the studied low-contrast mining-site environments.However, certain limitations may affect the generalizability of this study results. The dataset relied exclusively on UAV RGB images, with reference versions generated using machine learning techniques inspired by CLAHE. The absence of multimodal data (e.g., thermal or hyperspectral) restricts the model's adaptability to complex scenes. Furthermore, training images are mostly captured in homogeneous, low-contrast environments, limiting performance across other landscapes such as urban or forested areas.

Future research should therefore expand datasets with multimodal imagery to enhance generalization, incorporate more diverse environments to increase robustness, and perform more precise hyperparameter tuning to improve efficiency and stability. Finally, integrating the contrast enhancement model directly into photogrammetric workflows as a preprocessing step would facilitate its practical use in mining sites applications and maximize its impact on 3D terrain reconstruction quality.

### Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| ASPRS | American Society for Photogrammetry and Remote Sensing |
| C2C | Cloud-to-Cloud |
| CLAHE | Contrast Limited Adaptive Histogram Equalization |
| CNN | Convolutional Neural Network |
| DL | Deep Learning |
| DTM | Digital Terrain Model |
| GCP | Ground Control Point |
| ML | Machine Learning |
| MSE | Mean Squared Error |
| PSNR | Peak Signal-to-Noise Ratio |
| SSIM | Structural Similarity Index Measure |

| | |
|---|---|
| DTM | Digital Terrain Model |
| GCP | Ground Control Point |
| ML | Machine Learning |
| MSE | Mean Squared Error |
| PSNR | Peak Signal-to-Noise Ratio |
| SSIM | Structural Similarity Index Measure |

### Author Contributions

Khaoula ABKARI & Samira BEN AHMED: Writing - original draft, Writing - review & editing, Methodology, Conceptualisation, Formal analysis, Validation, Supervision, Project administration.
Sara AIT-LAMALLAM: Writing - review & editing, Methodology, Conceptualisation, Formal analysis, Validation, Supervision, Project administration.
Souhail KELLOUCH: Conceptualisation, Validation, Supervision, Project administration, Data curation.

### Funding
This research received no external funding
Data Availability Statement
Data not available - participant consent.

### Acknowledgements
The authors thank the Moroccan company AXIGEO for their warm welcome, technical support and supervision.

### Conflicts of Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 6. REFERENCES

Al Najjar, Y. (2024). Comparative analysis of image quality assessment metrics: MSE, PSNR, SSIM and FSIM. International Journal of Science and Research, 13(3), 110–114. https://doi.org/10.21275/SR24302013533

Anaconda. (2025). Anaconda: Open-source distribution for Python and data science. https://www.anaconda.com/

ArcGIS Pro. (2025). Desktop GIS software for mapping and spatial analysis. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Barron, J. T. (2017). Continuously differentiable exponential linear units. arXiv. https://doi.org/10.48550/arXiv.1704.07483

Cai, J., Gu, S., & Zhang, L. (2018). Learning a deep single-image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing, 27(4), 2049–2062. https://doi.org/10.1109/TIP.2018.2794218

Campos, G. F. C., Mastelini, S. M., Aguiar, G. J., Mantovani, R. G., Melo, L. F. D., & Barbon, S. (2019). Machine learning hyperparameter selection for CLAHE. EURASIP Journal on Image and Video Processing, 2019(1), 59. https://doi.org/10.1186/s13640-019-0445-4

Chen, S., Yuan, X., Yuan, W., & Cai, Y. (2016). Poor textural image matching based on graph theory. ISPRS Archives, XLI-B3, 741–747. https://doi.org/10.5194/isprsarchives-XLI-B3-741-2016

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of KDD, 785–794. https://doi.org/10.1145/2939672.2939785

Choi, O., Choi, J., Kim, N., & Lee, M. C. (2020). Combustion instability monitoring using deep learning. Electronics, 9(5), 848. https://doi.org/10.3390/electronics9050848

CloudCompare. (2025). 3D point cloud and mesh processing software. https://www.cloudcompare.org/

De Camargo, T., Schirrmann, M., Landwehr, N., Dammer, K.-H., & Pflanz, M. (2021). Optimized deep learning model for UAV weed mapping. Remote Sensing, 13(9), 1704. https://doi.org/10.3390/rs13091704

DJI. (2025). Mavic 2 product information. https://www.dji.com/be/mavic-2/info

Ehab, W., Huang, L., & Li, Y. (2024). UNet and variants for medical image segmentation. International Journal of Network Dynamics and Intelligence, 100009. https://doi.org/10.53941/ijndi.2024.100009

Gaou, B., Idrissi, A., Ait-Lamallam, S., & Kellouch, S. (2025). Good practices for 3D modeling of low-contrast areas using drone photogrammetry. In Sustainable Data Management (pp. 357–368). https://doi.org/10.1007/978-3-031-83915-3_28

Gindraux, S., Boesch, R., & Farinotti, D. (2017). Accuracy assessment of UAV-derived DSMs on glaciers. Remote Sensing, 9(2), 186. https://doi.org/10.3390/rs9020186

Gomes, D. M. (2008). Contrast enhancement in digital imaging using histogram equalization (PhD thesis). Université Paris-Est. https://theses.hal.science/tel-00470545v1

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light enhancement. arXiv. https://doi.org/10.48550/arXiv.2001.06826

Hai, J., Hao, Y., Zou, F., Lin, F., & Han, S. (2023). Advanced RetinexNet for low-light enhancement. Signal Processing: Image Communication, 112, 116916. https://doi.org/10.1016/j.image.2022.116916

Harichandana, M., Sowmya, V., Sajithvariyar, V. V., & Sivanpillai, R. (2020). Comparison of image enhancement techniques for post-flood images. ISPRS Archives, XLIV-M-2, 45–50. https://doi.org/10.5194/isprs-archives-XLIV-M-2-2020-45-2020

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980

Kryjak, T., Blachut, K., Szolc, H., & Wasala, M. (2022). Real-time CLAHE implementation for 4K video. Electronics, 11(14), 2248. https://doi.org/10.3390/electronics11142248

Lu, B., Zhu, J., Ge, Y., Chen, Q., Wen, Z., Liu, G., & Li, L. (2023). Automated volume estimation using terrestrial laser scanning. Remote Sensing, 15(18), 4604. https://doi.org/10.3390/rs15184604

Majidizadeh, A., Hasani, H., & Jafari, M. (2023). Semantic segmentation of UAV images using U-Net. ISPRS Annals, X-4/W1-2022, 451–457. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-451-2023

Marmolin, H. (1986). Subjective MSE measures. IEEE Transactions on Systems, Man, and Cybernetics, 16(3), 486–489. https://doi.org/10.1109/TSMC.1986.4308985

Matsimbe, J., Mdolo, W., Kapachika, C., Musonda, I., & Dinka, M. (2022). Drone vs. traditional volumetric computation in open-pit mining. Frontiers in Built Environment, 8, 1037487. https://doi.org/10.3389/fbuil.2022.1037487

Motayyeb, S., Fakhri, S. A., Varshosaz, M., & Pirasteh, S. (2022). Enhancing contrast to improve UAV photogrammetry accuracy. ISPRS Archives, XLIII-B1-2022, 389–398. https://doi.org/10.5194/isprs-archives-XLIII-B1-2022-389-2022

Mousavi, V., Varshosaz, M., & Remondino, F. (2021). Evaluating tie-point distribution in UAV photogrammetry. ISPRS Archives, XLIII-B2-2021, 39–46.https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-39-2021

Nilsson, J., & Akenine-Möller, T. (2020). Understanding SSIM. arXiv. https://doi.org/10.48550/arXiv.2006.13846

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An imperative deep learning library. arXiv. https://doi.org/10.48550/arXiv.1912.01703

Photogrammetric Engineering & Remote Sensing. (2015). ASPRS positional accuracy standards. Photogrammetric Engineering & Remote Sensing, 81(3), 1–26. https://doi.org/10.14358/PERS.81.3.A1-A26

PIX4Dmapper. (2025). Photogrammetry software for drone mapping. https://www.pix4d.com/product/pix4dmapper-photogrammetry-software

Pizer, S. M., Austin, J. D., Perry, J. R., Safrit, H. D., & Zimmerman, J. B. (1986). Adaptive histogram equalization for medical images. In Application of Optical Instrumentation in Medicine XIV (p. 242). https://doi.org/10.1117/12.975399

QGIS. (2025). QGIS: Spatial without compromise. https://qgis.org/

Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., & Shoyaib, M. (2016). Adaptive gamma correction for image enhancement. EURASIP Journal on Image and Video Processing, 2016(1), 35. https://doi.org/10.1186/s13640-016-0138-1

Rath, S., Priyadarshini, S. B. B., Sahu, P. K., Panda, M., Patel, D. K., Jagadev, N., & Patra, N. (2024). Enhancing visual content quality using MIRNet-V2 and autoencoders. Journal of Theoretical and Applied Information Technology, 102(15), 5783–5793.

DJI. (2025). Support for Mavic 2. https://www.dji.com/be/support/product/photo

Tanchenko, A. (2014). Visual-PSNR measure of image quality. Journal of Visual Communication and Image Representation, 25(5), 874–878. https://doi.org/10.1016/j.jvcir.2014.01.008

Tao, Q., Liu, R., Li, X., Gao, T., Chen, Y., Xiao, Y., He, H., & Wei, Y. (2025). Monitoring 3D surface deformation using SBAS-InSAR and GNSS. Scientific Reports, 15(1), 2853. https://doi.org/10.1038/s41598-025-87087-4

Touhami, M., Ur Rehman, Z., Ahmad Fauzi, M. F., & Mansor, S. B. (2024). Comparison of conventional and U-Net-based histopathology enhancement. In ICSIPA 2024 (pp. 1–6). https://doi.org/10.1109/ICSIPA62061.2024.10686178
Visual Studio Code. (2025). Code editing platform. https://code.visualstudio.com/

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep Retinex decomposition for low-light enhancement. arXiv. https://doi.org/10.48550/arXiv.1808.04560

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). Structure-from-Motion photogrammetry for geoscience. Geomorphology, 179, 300–314. https://doi.org/10.1016/j.geomorph.2012.08.021

Xu, Y., Zhang, X., & Liu, J. (2020). Adaptive local contrast enhancement for low-visibility aerial images. In MIPPR 2019 (p. 16). https://doi.org/10.1117/12.2538091

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., & Shao, L. (2020). Learning enriched features for real image restoration. arXiv. https://doi.org/10.48550/arXiv.2003.06792

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., & Shao, L. (2023). Learning enriched features for fast image restoration. IEEE TPAMI, 45(2), 1934–1948. https://doi.org/10.1109/TPAMI.2022.3167175

Zhang, L., Ge, D., Guo, X., Liu, B., Li, M., & Wang, Y. (2020). InSAR monitoring of mining-induced deformation. In IAHS Proceedings (pp. 237–240). https://doi.org/10.5194/piahs-382-237-2020

# EXAMINING ALBANIAN MUNICIPAL FINANCE THROUGH A DIGITAL AUDIT TRIAGE USING BENFORD'S LAW

***Steven John Asllani*** [1],
*1 University Metropolitan Tirana, Albania*

## Abstract

This article develops and demonstrates a Benford-based audit triage framework for municipal treasury payments in Albania. The study addresses a practical public-sector audit problem: local governments generate payment populations that are too large for full manual verification, while audit teams still require a transparent method for prioritizing transactions, years, expenditure classes and value bands for further examination. Using 2,868 treasury payment transactions from the Municipality of Saranda for 2021–2025, the article applies first-digit and first-two-digit Benford tests, mean absolute deviation (MAD) and chi-square goodness of +fit statistics after cleaning the payment population and separating regulated or administratively determined streams from discretionary expenditure. The results show that pooled testing is misleading because regulated streams produce structural nonconformity: the regulated stream first-digit MAD of 0.028173 substantially exceeds the discretionary population first-digit MAD of 0.008248. In the discretionary population, first-digit conformity is broadly acceptable, whereas first-two-digit testing (MAD = 0.004669) reveals localised deviations concentrated in 2024, service-related expenditure, threshold-adjacent value bands and recurring same-amount patterns. Chi-square tests broadly confirm the MAD findings, with the discretionary population producing lower chi-square statistics than the pooled or regulated populations, though formal significance is interpreted cautiously given sample-size sensitivity. The contribution of the article is not to treat digit deviations as proof of fraud, but to convert them into an audit workflow: population design, segmentation, anomaly ranking, directed sample selection and documentary follow-up through contract, invoice, procurement and approval testing. The findings support a cautious but operationally useful role for Benford analysis in Albanian municipal auditing when it is governed by explicit population exclusion rules, minimum-sample safeguards, professional judgement and corroborating evidence.

**Keywords:** Benford's Law; audit analytics; municipal treasury payments; public-sector auditing; fraud risk; Albania

## 1. INTRODUCTION

Municipal audit work is increasingly shaped by the tension between large digital transaction populations and limited audit resources. Local governments process payments for services, maintenance, capital works, routine administration and legally regulated transfers. A traditional audit approach based only on judgmental sampling or small document checks can miss clustered irregularities, repeated payment patterns and threshold-related behaviors. Audit teams therefore need screening procedures that are inexpensive, repeatable and capable of converting large payment populations into a manageable set of higher-risk audit leads. Benford's Law is one such procedure. It predicts that the leading digits of many naturally generated numerical datasets follow a logarithmic rather than uniform distribution. In audit analytics, a departure from the expected distribution can indicate artificial number construction, unusual concentration, repeated values, transaction splitting or other patterns that deserve follow-up. However, modern audit literature is unequivocal that Benford deviations cannot be read as proof of fraud. They are risk signals, not verdicts. This distinction is particularly important in the public sector, where values are often shaped by tariffs, legal limits, procurement thresholds and recurring administrative payments all of which can produce non-Benford digit distributions without any misconduct. This article applies that cautious interpretation to municipal treasury payments in Albania. The empirical setting is the Municipality of Saranda, using payment transactions covering 2021–2025. The study is designed as an audit triage exercise. It asks whether digit analysis can help auditors rank payment segments for further testing, and under what conditions such ranking is defensible. The core argument is that Benford analysis becomes useful only after careful population design, segmentation of regulated and discretionary expenditure, and deliberate translation of statistical anomalies into substantive audit procedures. The International Standard on Auditing 240 (IAASB, 2024) provides the professional framework within which this analytical approach sits: the auditor's

*Corresponding author: Steven John Asllaniai, steven.asllani34@gmail.com

responsibility is to maintain professional skepticism and respond to fraud risk indicators with appropriate procedures, and a structured Benford screen is one tool for identifying which segments warrant that closer attention. The article differs from a simple conformity test in two important ways. It does not ask merely whether the dataset follows Benford's Law, it asks how deviations can be operationalized within a municipal audit workflow. Treats the first-two-digit test as the central audit tool, because the first-digit test is too coarse to detect the localized concentrations, threshold-adjacent value bands, recurring vendor amount pairs, service-contract clusters that matter most for procurement and approval controls.

The study addresses three research questions: (1) under what conditions are municipal treasury payments suitable for Benford testing; (2) which segments generate the most audit-relevant deviations; and (3) how can digit anomalies be converted into risk-ranked sample selection and follow-up procedures, Section 2 reviews the relevant literature. Section 3 develops the conceptual framework. Section 4 describes the data and methodology. Section 5 presents the empirical results. Section 6 discusses the findings and their implications. Section 7 outlines the workflow for audit practice. Section 8 concludes.

## 2. LITERATURE REVIEW

The intellectual origin of Benford's Law is usually traced to Newcomb (1881), who observed that lower-numbered pages in logarithm tables were used more heavily than higher-numbered pages. Benford (1938) later tested the phenomenon across a wider range of empirical datasets and formalised the expected logarithmic distribution. For the first significant digit d, the probability is $P(d) = \log_{10}(1 + 1/d)$, where d ranges from 1 to 9. The logic extends to first-two-digit combinations from 10 to 99, allowing more granular testing of leading digit pairs. The accounting and audit literature transformed Benford's Law from a mathematical regularity into a practical analytical procedure. Nigrini and Mittermaier (1997), Drake and Nigrini (2000), Nigrini (2001, 2012) and Durtschi, Hillison and Pacini (2004) established the use of digital analysis for fraud screening, audit planning and transaction review. Their contribution was not only statistical: it showed that digit patterns can direct audit attention toward unusual records, suppliers or periods when the underlying population is appropriate. Early Albanian applications of these methods, including Asllani and Naco (2013), demonstrated that Benford-based screening can identify suspicious accounting patterns in Albanian institutional data providing a foundation on which the present study builds by extending the method to municipal treasury payments and embedding it in an explicit audit triage workflow. Later literature has become more disciplined and more cautious. Nigrini (2017) reviews Benford-based audit sampling while emphasising population suitability, sample-size issues and methodological limits. Druica, Oancea and Valsan (2018) show that digit analysis may produce ambiguous or conflicting results depending on how the sample is decomposed and which conformity metric is used. Nigrini (2019) links fraud-related number patterns to duplicates, round amounts and values near control thresholds, but also stresses the need to refine tests and manage false positives. The most recent research moves beyond single-test Benford applications. Leonov et al. (2022) demonstrate integrated Benford testing for corporate fraud detection. Le and Mantelaers (2024) argue for a broader framework that combines Benford analysis with additional statistical and machine-learning procedures. Cano-Rodriguez, Nunez-Nickel and Liceran-Gutierrez (2025) provide an important critical counterweight by showing that divergence from Benford's Law is not a stable stand-alone measure of financial statement accuracy: digit patterns can arise for structural reasons that have no connection to irregularity, and conclusions drawn from Benford analysis alone are not reliable indicators of misstatement. The implication for public-sector audit is clear, Benford analysis should be embedded in a broader evidentiary process, not used as a self-sufficient diagnostic.

Municipal finance is a particularly demanding setting for this literature. Public payments are not always naturally generated market values. They can include regulated transfers, standard service contracts, fixed fees, repeated administrative amounts and threshold-shaped procurement behaviour. These features create non-Benford patterns without any misconduct. Responsible use in municipal auditing therefore requires the analyst to separate structurally unsuitable streams before testing, interpret deviations in their institutional context and link each red flag to specific documentary corroboration rather than to an allegation.

## 3. CONCEPTUAL FRAMEWORK: FROM DIGIT SIGNAL TO AUDIT RESPONSE

The conceptual framework used in this article treats Benford analysis as the first stage of an audit triage cycle. The cycle begins with the construction of a valid transaction population one that excludes non-discretionary values that would contaminate the digit distribution. It then proceeds to first-digit and first-two-digit screening, localization of anomalies by subgroup, contextual interpretation of the anomaly in relation to municipal payment structures, and selection of documentary follow-up procedures. The final output is not a fraud conclusion but a documented, evidence-linked decision about where audit work should be concentrated. The framework has three interpretive layers. The statistical layer

measures whether observed leading-digit frequencies differ from Benford expectations, using MAD as the primary magnitude indicator and chi-square as a formal inferential check. The contextual layer asks whether the deviation is likely to be explained by lawful institutional structure fixed tariffs, recurring services, procurement-threshold rounding rather than by misconduct. The audit-response layer specifies which transactions or segments should be examined through invoices, contracts, procurement records, receiving documentation and approval trails, consistent with the fraud-response requirements of ISA 240 (IAASB, 2024). The framework also defines an interpretive boundary that governs the entire study. A digit deviation is treated as a meaningful audit lead only when it is persistent across multiple observations, concentrated in an audit-relevant segment rather than scattered, and capable of being tested against external documentary evidence. A deviation in a small value band justifies a targeted follow-up test but not a broad claim. A deviation in a regulated stream reflects structural nonconformity rather than fraud risk. This boundary is what distinguishes a responsible Benford application from an over-interpretation of statistical noise.

| Audit stage | Purpose and audit meaning |
|---|---|
| Population design | Exclude non-amount fields and separate regulated from discretionary streams. This is the most important preparatory step: without it, structural nonconformity in regulated payments will dominate the distribution and produce misleading risk signals. |
| Digit screening | Apply first-digit and first-two-digit tests; compute MAD and chi-square. Use MAD as the primary practical indicator of deviation magnitude; treat chi-square significance as corroborating rather than conclusive given its sensitivity to sample size. |
| Segmentation | Break down results by year, expenditure category, value band and recurring-amount pattern to localise the source of elevated deviation. Population-level conformity may conceal segment-level anomalies that are the true audit leads. |
| Audit response | Rank hotspots by deviation magnitude, observation count and audit relevance. Select directed samples from high-priority segments combined with stratified samples from the broader discretionary population. |
| Corroboration | Vouch invoices, contracts, approval records and procurement files for sampled transactions. No finding is reportable without documentary evidence. The digit screen identifies where to look; corroboration determines what was found. |

**Table 1:** *Audit triage logic: stages, purposes and audit meaning*

# 4. DATA AND METHODOLOGY

## *4.1. Case selection and data*

The empirical case is the Municipality of Saranda, a mid-sized Albanian coastal municipality. The case is appropriate for this study for three reasons. The transaction volume (2,868 treasury payments over five years) is sufficient for meaningful digit testing while remaining manageable for case-level analysis. The payment structure is heterogeneous: Saranda processes both regulated transfers and discretionary expenditure across multiple economic classifications, making it a realistic test of the population-segmentation logic. Albanian municipalities of this scale are representative of the environment in which any resulting audit workflow would be deployed, supporting practical transferability.
The dataset contains 2,868 treasury payment transactions for 2021–2025. The unit of analysis is the individual monetary payment amount. Non-amount identifiers, administrative reference codes and other non-numeric fields are excluded from digit testing before analysis begins. The raw payment population is then divided into three analytical populations. P0 refers to all valid payments after basic data cleaning (N = 2,868). Regulated or administratively determined streams, comprising fixed transfers, salary-related payments and tariff-governed items are extracted and analysed separately as a diagnostic population (N = 1,442), because their fixed or formulaic values are not expected to conform to Benford's Law regardless of the integrity of the underlying transactions. P1 refers to discretionary expenditure, including goods, services, maintenance, transport and capital expenditure (N = 1,426). P1 is the primary Benford-testing population because it is the most likely to contain naturally variable monetary values that the Benford distribution presupposes.

## *4.2. Tests and decision logic*

The analysis uses two complementary digit tests applied to each analytical population. The first-digit test compares observed frequencies for leading digits 1 to 9 with the expected Benford probabilities. The first-two-digit test compares observed frequencies for leading digit pairs 10 to 99 with expected probabilities. The first-digit test provides a broad conformity signal suitable for population-level screening. The first-two-digit test is used as the primary audit tool because it is more sensitive to localised concentrations specific value clusters, threshold bands or vendor patterns, that cannot be detected at the first-digit level Conformity is evaluated using two statistics. Mean absolute deviation (MAD) is the primary practical measure because it expresses the average magnitude of deviation between observed and expected digit frequencies in proportional terms, making it directly interpretable without reference to sample size.

Chi-square is retained as a formal goodness-of-fit test. It is computed for all populations and segments but is interpreted cautiously because test significance is sensitive to N: large samples will reject Benford conformity even for trivially small deviations, while small subsamples may fail to detect genuine concentrations. Where chi-square results are reported, they are presented alongside MAD to give a complete picture. The audit interpretation follows a risk-based rule: a deviation is treated as an audit lead when it is concentrated, persistent across multiple periods or categories, operationally meaningful in the context of municipal payment structures, and capable of documentary verification

## 5. RESULTS

### *5.1. Population-level conformity*

Table 2 summarizes the conformity results for the three main analytical populations. The pooled result for all payments (P0) shows a first-digit MAD of 0.014863. Under the conventional MAD benchmarks for first-digit testing (Nigrini, 2012: close conformity < 0.006; acceptable conformity 0.006–0.012; marginal 0.012–0.015; nonconforming > 0.015), P0 falls at the boundary of marginal and nonconforming. This result, taken in isolation, might suggest a problematic population. In fact, it reflects the mixing of structurally different payment streams. The regulated stream confirms this: its first-digit MAD of 0.028173 indicates strong structural nonconformity. This is expected and carries no fraud implication, regulated payments include fixed transfers and tariff-based items whose leading digits are institutionally determined rather than naturally generated. Including these in a pooled Benford test contaminates the distribution and produces a misleading audit signal. The chi-square statistic for P0 ($\chi^2 = 31.4$, df = 8, $p < .001$) and for the regulated stream ($\chi^2 = 58.2$, df = 8, $p < .001$) both confirm statistically significant departure from Benford, but the regulated-stream result reflects institutional structure, not risk. After isolating the discretionary population P1, the picture changes substantially. P1 shows first-digit MAD of 0.008248, falling within the acceptable conformity range. The chi-square for P1 first-digit testing ($\chi^2 = 9.7$, df = 8, $p = .286$) is non-significant, supporting broad first-digit acceptability. However, the first-two-digit MAD for P1 is 0.004669. While this value appears smaller in absolute terms, first-two-digit benchmarks are lower than first-digit benchmarks because the expected proportions are smaller (90 possible pairs vs 9 digits): a first-two-digit MAD above 0.0040 typically indicates noteworthy localized deviation. The P1 first-two-digit chi-square ($\chi^2 = 112.6$, df = 89, $p = .046$) is marginally significant, confirming that pockets of meaningful deviation exist within an otherwise broadly conforming discretionary population.

| Population | First-digit MAD | $\chi^2$ (df = 8) | p-value | Audit interpretation |
|---|---|---|---|---|
| P0: all valid payments (N = 2,868) | 0.014863 | 31.4 | < .001 | Pooled result is distorted by regulated streams; do not use for audit targeting. |
| Regulated streams (N = 1,442) | 0.028173 | 58.2 | < .001 | Strong structural nonconformity; reflects institutional determination, not fraud risk; analyse separately. |
| P1: discretionary payments (N = 1,426) | 0.008248 (1st digit) 0.004669 (1st–2nd digit) | 9.7 (1st digit) 112.6* (1st–2nd, df = 89) | p = .286 p = .046 | Acceptable first-digit conformity; localised first-two-digit deviations remain; P1 is the suitable primary triage population. |

**Table 2:** *Summarizes of the conformity results*

*Author's notes* MAD benchmark interpretation follows Nigrini (2012): close conformity < 0.006 (first-digit scale); acceptable 0.006–0.012; marginal 0.012–0.015; nonconforming > 0.015.*

### *5.2. Localized deviations*

Within P1, the first-two-digit test identifies the most useful risk signals. Excess frequencies are concentrated in a limited set of leading pairs, especially digit pairs 20, 11, 85, 30 and 76. These pairs are not interpreted in isolation; they serve as pointers to the transactions, years and economic categories that generate the concentrated frequencies. The analysis below traces these concentrations through three segmentation lenses: time, economic category and value band.
Time segmentation shows that first-two-digit deviation persists across the full study period and peaks in 2024. Annual first-two-digit MAD values are: 0.00654 (2021), 0.00598 (2022), 0.00659 (2023), 0.00788 (2024) and 0.00699 (2025). The 2024 value is the highest in the series and represents a meaningful elevation relative to the surrounding years. 2024 is therefore the most important temporal hotspot for audit planning: it is the year where the concentration of localized digit anomalies is highest and where directed sampling should begin. Economic classification adds further precision. Table 3 summarizes the category-level results. The Services category is the most decision-relevant because it combines a large number of observations (N = 619, representing 43.4% of P1) with elevated first-two-digit deviation (MAD = 0.006802). This combination, size and deviation together, makes Services the highest-priority category for follow-up. The Other or Unclassified category shows the highest MAD of any category (0.008817) and accounts for 187

transactions, or 13.1% of P1. The audit interpretation for this category differs from Services: the elevated MAD may reflect coding and classification problems rather than payment irregularity, and the appropriate first step is to recode a representative sample of records and test data-quality controls before drawing any payment-level inference. Goods/Materials (N=312) and Maintenance/Transport (N=308) show elevated MAD values for their sizes but their smaller observation counts require more cautious claims and narrower targeted samples.

Value-band analysis also identifies threshold-related signals. Bands near administrative or procurement approval thresholds display elevated deviation and heavy rounding: strong concentrations of values ending in 00 are present in threshold-adjacent bands. This pattern is audit-relevant because threshold behavior is consistent with either legitimate administrative routines, standard contract amounts, recurring service fees or possible payment splitting intended to keep individual transactions below approval or procurement requirements. The appropriate response is not accusation but targeted procurement and approval testing for transactions in these bands.
Recurring same-vendor same-amount patterns explain part of the first-two-digit deviation. Within P1, repeated transactions where the same vendor code appears with the same payment amount account for a meaningful proportion of the excess frequencies in digit pairs 20, 30 and 76. These recurrences are not in themselves irregular, they may reflect legitimate recurring service contracts but they raise three specific audit questions: Is each payment supported by a distinct invoice reflecting actual service delivery? Is the contract basis for the recurring series properly documented and within the approved procurement method? Are any pairs or series duplicates that were processed more than once in error? These are the questions that the digit screen converts into audit action, consistent with the corroboration requirement set out in the framework.

| Category | N | % of P1 | 1st–2nd MAD | Recommended audit response |
|---|---|---|---|---|
| Services | 619 | 43.4% | 0.006802 | Vouch contract basis, invoice support and service acceptance documentation for sampled transactions. Prioritise 2024. Check approval chain for high-value service payments. |
| Other / Unclassified | 187 | 13.1% | 0.008817 | Recode a sample of records; test data-quality controls. Interpret digit signal only after classification is validated. |
| Goods / Materials | 312 | 21.9% | 0.007115 | Test procurement method for sampled items; verify goods receipt against invoice and delivery documentation. |
| Maintenance / Transport | 308 | 21.6% | 0.006940 | Check contract or framework agreement basis; verify that individual payments align with approved scope and rates. |
| Total P1 | 1,426 | 100% | 0.004669 (pooled) | Stratified sample across categories; directed sample from Services and Other for 2024 priority period. |

**Table 3**: *Economic category breakdown*

*Author's notes. Categories with N < 100 are not suitable for standalone Benford conclusions and are treated as audit leads only in combination with category-level risk factors.*

## 6. DISCUSSION

The findings support the central proposition of the article: Benford analysis is useful for municipal audit planning only when it is applied selectively and interpreted through a wider audit framework. The Saranda evidence illustrates both the risk of mechanical application and the payoff from disciplined segmentation. The pooled result (P0 first-digit MAD = 0.014863) would appear alarming in isolation, yet it is entirely explained by the structural nonconformity of regulated streams. Once discretionary expenditure is isolated, broad first-digit acceptability is restored, while first-two-digit testing identifies concentrated deviations that can be translated into specific audit action.
This result is directly consistent with the established literature. It supports Nigrini's argument that Benford analysis assists audit sampling and planning when the auditee data are appropriate and demonstrates the corollary that inappropriately pooled populations undermine that assistance. It confirms the caution of Durtschi et al. and Druica et al., because the Saranda pooled result is precisely the kind of misleading output they warned against. It aligns with the beyond-Benford perspective of Le and Mantelaers (2024), because practical value emerges only after segmentation, contextual interpretation and follow-up procedures are added to the statistical screen. And it is consistent with the critical finding of Cano-Rodriguez et al. (2025) that divergence from Benford's Law should not be treated as a stand-alone measure of financial irregularity a caution that is built into every layer of the triage framework used here. The first-two-digit analysis is the paper's strongest empirical

contribution. In the discretionary population, first-digit results indicate broad acceptability but first-two-digit tests reveal audit-relevant clusters in specific years, categories and value bands. This matters because irregularities and control weaknesses in public-sector payments rarely affect the leading digit alone. They typically manifest as value bunching around thresholds, repeated contract amounts, standardised invoices or recurring supplier patterns all of which produce localised first-two-digit concentrations that a first-digit comparison will not detect.

The governance message is the second key finding. A Benford workflow improves audit transparency by creating a documented, replicable basis for sample selection. The auditor can demonstrate that a specific year, category, digit pair or value band displayed elevated deviation and therefore justified the allocation of substantive testing resources. This structures professional judgement rather than replacing it, and creates an audit trail from data screen to audit procedure that is both internally coherent and externally defensible. The third message concerns limits. Vendor identifier missingness, heavy rounding and recurring payments reduce the precision with which digit deviations can be attributed to specific transactions or suppliers. The recurring same-vendor same-amount patterns identified in the Saranda dataset explain part of the first-two-digit deviation and partially account for the 2024 peak. These patterns do not invalidate the method; they redirect the audit question from "is there a digit deviation?" to "is the recurring series contractually justified, properly approved, supported by evidence of service delivery and free from duplicate-payment risk?" That is a more precise and more useful question. In comparative perspective, the Saranda findings are broadly consistent with what Benford studies in other small municipal and transitional-economy contexts have reported. Asllani and Naco (2013) found that Albanian accounting datasets exhibit identifiable digit patterns amenable to Benford screening, and the present study extends that observation to the treasury payment level with an explicit triage operationalisation. Leonov et al. (2022), working in a corporate context, found that integrated multi-test Benford approaches produce more reliable signals than single tests a finding that the current two-test (first-digit plus first-two-digit) design reflects. The Saranda results do not establish that Albanian municipal finance is more or less irregular than comparable settings; they establish that the triage method works as intended and produces actionable differentiation within a real payment population.

Returning to the first research question under what conditions are municipal treasury payments suitable for Benford testing the evidence from Saranda supports the following synthesis. A municipal payment population is suitable for Benford-based audit triage when five conditions are met: (1) the population is large enough to support stable digit frequencies, with a minimum of approximately 500 transactions in the primary testing population and at least 100 per sub-segment before drawing segment-level conclusions; (2) regulated, tariff-based and administratively determined streams are separated prior to testing and analysed independently with an explicit note that their nonconformity carries no fraud implication; (3) both first-digit and first-two-digit tests are run, since first-digit testing alone is insufficient to detect the localised concentrations that are most audit-relevant; (4) deviation results are segmented by year, economic category and value band before any audit decision is made, because population-level conformity can mask segment-level anomalies and vice versa; and (5) every statistical deviation that is elevated to an audit lead is matched to a specific documentary follow-up procedure invoice, contract, procurement file or approval record rather than treated as a self-sufficient finding. When these five conditions are not met, Benford testing should be deferred or its output treated as exploratory rather than as an audit planning tool.

## 7. IMPLICATIONS FOR AUDIT PRACTICE

For Albanian public-sector audit practice, the article proposes a five-step workflow that translates the conceptual framework and the Saranda results into repeatable audit procedure. The workflow is designed to be applicable by external audit institutions, municipal internal audit units and financial inspection teams operating with standard spreadsheet or data-analysis tools.

Step one is population definition. Before any digit test is run, the auditor must document which payment streams are included in the primary population, which are excluded, and the reason for each exclusion. Regulated, tariff-based or routine administrative streams should either be removed from the primary Benford population or placed in a separate diagnostic population with the explicit notation that their digit patterns are institutionally determined. This step is the most important governance safeguard in the entire workflow skipping it produces the pooled-testing distortion demonstrated in Section 5. Step two is dual-test execution. Both first-digit and first-two-digit tests should be run on the primary discretionary population. MAD is the primary reported statistic; chi-square is computed and reported as a formal corroborating check, with its sensitivity to sample size explicitly acknowledged in any working paper or report.

Step three is segmentation. Results should be broken down by year, economic classification, value band and recurring-amount pattern before any ranking decision is made. The objective is to identify not just

that deviation exists but where it is concentrated. A single year, category or digit pair that accounts for a disproportionate share of the deviation is a more actionable lead than an evenly distributed deviation across the whole population. Step four is ranked sample selection. Segments are ranked by deviation magnitude, observation count, rounding intensity and operational audit relevance. Directed samples should be drawn from the highest-ranked segments; stratified samples should cover the broader discretionary population. The combination ensures that the most anomalous areas receive priority attention without leaving the rest of the population unsampled. Step five is documentary corroboration. Selected transactions are tested through invoice vouching, contract review, procurement method verification, receiving documentation and approval trail testing, consistent with ISA 240 (IAASB, 2024) requirements for responding to identified fraud risk indicators. No conclusion whether positive or negative should be drawn from the digit screen alone.

The workflow also highlights data governance needs that are relevant beyond the immediate audit engagement. Consistent vendor identifiers, cleaner economic classification, duplicate-payment controls and standardised payment descriptions would substantially improve the reliability and precision of audit analytics. Without these fields, digit analysis can still identify hotspots, but vendor-level and supplier-pattern conclusions remain weaker than they would be with clean reference data. The method therefore has a dual function: it screens for audit risk and exposes weaknesses in the data infrastructure needed for modern public-sector financial oversight.

| **Safeguard** | **Practical rule and rationale** |
|---|---|
| Explicit population definition | Document all inclusions and exclusions with reasons before running any test. Undocumented exclusions undermine the defensibility of the triage output. |
| Separate regulated streams | Never use regulated or tariff-based streams as direct fraud screens. Their structural nonconformity is institutionally determined and carries no inferential value for misconduct. |
| Minimum-N discipline | Treat sub-segments with fewer than 100 transactions as directional leads only, not as standalone conclusions. Report the N alongside every MAD value so readers can calibrate confidence appropriately. |
| Dual-test reporting | Always report both first-digit and first-two-digit MAD. Report chi-square with df and p-value, noting its sample-size sensitivity. Never rely on first-digit results alone. |
| Supervisor review | Have a second auditor review the population exclusion decisions, segmentation choices and shortlist logic before any transactions are selected for follow-up. This reduces both analytical error and the risk of selective attention. |
| Mandatory corroboration | No allegation, finding or management letter observation should be based on digit patterns alone. Every elevated segment must be followed up with invoice, contract, approval record or procurement file evidence before any conclusion is reached. |

**Table 4**: *Minimum governance safeguards*

## 8. CONCLUSION

This article examined whether Benford's Law can support municipal audit planning in Albania through a case study of Saranda treasury payments for 2021–2025. The evidence shows that Benford screening is valuable but only under strict methodological conditions. Pooled testing of all payments is misleading because regulated streams generate structural nonconformity that has no fraud implication. The discretionary payment population is more suitable: it shows broad first-digit acceptability, confirmed by non-significant chi-square results, but meaningful first-two-digit deviations marginally significant at the population level and concentrated in 2024, the Services category and threshold-adjacent value bands that point to specific audit priorities.

The main contribution is procedural. The study converts digit analysis into a five-stage audit triage framework that moves from population design through anomaly ranking to documentary follow-up,

with explicit governance safeguards at each stage. This is a substantively stronger contribution than reporting whether a dataset conforms to Benford's Law. It shows how a municipal auditor can use digital analysis to justify sample selection, strengthen fraud-risk planning under ISA 240, and document the chain from statistical signal to audit procedure in a way that is both internally coherent and professionally defensible.

The study also sets clear interpretive limits. Benford deviations are not proof of fraud, corruption or misstatement. They are structured indicators of where the auditor should ask better questions and apply more targeted procedures. The Saranda findings particularly the recurring same-vendor same-amount patterns, the 2024 temporal peak and the Services category concentration exemplify the kind of operationally actionable output that a responsibly applied digit screen can produce.

Three directions for future research follow naturally from this work. First, extending the framework to other Albanian municipalities particularly those of different sizes and sectoral compositions, such as Shkodra or Elbasan would test whether the Saranda patterns are idiosyncratic or characteristic of Albanian municipal payment structures more broadly. Second, combining treasury payment data with procurement metadata vendor registration details, contract awards, procurement method classifications and approval levels would substantially improve the precision of vendor-level and threshold-related conclusions, addressing the data-governance limitation identified in this study. Third, and most consequentially, testing whether Benford hotspots correlate with subsequent audit findings that is, whether segments ranked as high-priority by digit triage actually produce more substantive findings when examined would provide the first empirical validation of whether digit-based triage improves audit efficiency relative to judgmental or random selection. Such a longitudinal study, drawing on the Albanian Supreme Audit Institution's (ALSAI) multi-year audit recommendation data, would determine whether digit-based triage can become a standard and evidence-validated component of public-sector audit analytics in Albania.

**Acknowledgements**

The author acknowledges the academic supervision received during the master's thesis research process at Metropolitan University. The dataset was obtained from publicly available treasury records of the Municipality of Saranda. The interpretations, analytical decisions and conclusions presented in this article are the sole responsibility of the author. No part of this article has been previously published or is under review elsewhere.

## 9. REFERENCES

Asllani, A., & Naco, M. (2013). Benford's law as an aid tool for detecting suspicious patterns in accounting data: Evidence from Albania. Albanian Academic Review on Accounting and Audit Analytics.

Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, 78(4), 551–572.

Cano-Rodriguez, M., Nunez-Nickel, M., & Liceran-Gutierrez, A. (2025). Divergence from Benford's law fails to measure financial statement accuracy. International Journal of Accounting Information Systems, 56, 100745.

Drake, P. D., & Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford's Law. Journal of Accounting Education, 18(2), 127–146.

Druica, E., Oancea, B., & Valsan, C. (2018). Benford's law and the limits of digit analysis. International Journal of Accounting Information Systems, 31, 75–82.

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of Forensic Accounting, 5, 17–34.

International Auditing and Assurance Standards Board. (2024). International Standard on Auditing 240: The auditor's responsibilities relating to fraud in an audit of financial statements. IFAC.

Le, L., & Mantelaers, E. (2024). Benford's Law and beyond: A framework for auditors. Maandblad voor Accountancy en Bedrijfseconomie, 98(7), 427–438.

Leonov, P. Y., Suyts, V. P., Norkina, A. N., & Sushkov, V. M. (2022). Integrated application of Benford's Law tests to detect corporate fraud. Procedia Computer Science, 213, 332–337.

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics, 4(1), 39–40.

Nigrini, M. J. (2001). Digital analysis using Benford's law: Tests and statistics for auditors. Global Audit Publications.

Nigrini, M. J. (2012). Benford's Law: Applications for forensic accounting, auditing, and fraud detection. Wiley.

Nigrini, M. J. (2017). Audit sampling using Benford's Law: A review of the literature with some new perspectives. Journal of Emerging Technologies in Accounting, 14(2), 29–46.

Nigrini, M. J. (2019). The patterns of the numbers used in occupational fraud schemes. Managerial Auditing Journal, 34(5), 606–626.

Nigrini, M. J. (2022). Using Benford's Law to reveal journal entry irregularities. Journal of Accountancy.
Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's Law as an aid in analytical procedures. Auditing: A Journal of Practice & Theory, 16(2), 52–67.

Tota, I., Aliaj, A., & Lamcja, D. (2016). Practical conditions and limits of Benford analysis in audit applications. Albanian Academic Review on Accounting and Audit Analytics.

# A MACHINE LEARNING APPROACH FOR PREDICTING LAW APPROVAL PROBABILITY: A CASE STUDY OF NORTH MACEDONIA

***Safije Sadiki Shaini*** [1] ***,Majlinda Fetaji*** [2]

*1 South East European University, Tetovo, North Macedonia*

## Abstract

Today, every parliamentary system generates large amounts of data, enabling predictive analytics within digital parliamentary systems. This study uses machine learning techniques to predict the probability of Law Approval (PLA) as a case study of the Assembly of the Republic of North Macedonia. A dataset was created from scratch, where legislative, procedural, political, and institutional data were collected from public sources covering 9 regular and technical government mandates from 2008 to 2026. Several supervised machine learning classifiers, including Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and XGBoost, were trained and evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. Models were trained using an 80/20 train-test split and Stratified 5-Fold Cross-Validation. To address class imbalance, SMOTETomek and class-weight balancing techniques were applied during model training. Random Forest achieved the best overall performance with an 80/20 train-test split, reaching an Accuracy of 0.847, an F1-score of 0.876, and a ROC-AUC value of 0.910. XGBoost and Gradient Boosting also achieved good classification results. The results show that machine learning techniques can be used for prediction, as in our case for PLA within electronic parliamentary systems.

**Keywords:** e-Parliament · Machine Learning · Probability of Law Ap- proval · Machine Learning Models · PLA

## 1. INTRODUCTION

Modern parliamentary systems generate large volumes of legislative and procedural data through digital parliamentary platforms, creating new opportunities for predictive analytics and intelligent decision support within e-Parliament systems (Fitsilis, Koryzis, & Schefbeck, 2022; Cavalieri et al., 2022). Contemporary legislative institutions increasingly rely on digital technologies to improve transparency, accountability, and institutional efficiency within parliamentary procedures (Reis, Santo, & Melão, 2019). Predicting legislative approval can support transparency, institutional planning, and evidence-based governance by helping parliamentary institutions better understand which characteristics are associated with successful legislative approval. outcomes Arslan (2025). Machine learning models are increasingly used in legal analytics and parliamentary prediction tasks because of their ability to detect hidden patterns in large datasets and generate reliable predictions (Sulea et al., 2017).

Recent studies demonstrate that ML techniques can successfully be applied in parliamentary and political contexts, including voting prediction, sentiment analysis of parliamentary debates, legislative speech analysis, and forecasting whether a bill will become law (Goldblatt & O'Neil, 2012). Machine learning techniques were used to predict the likelihood of legislative approval based on textual and procedural information extracted from parliamentary bills (Nay, 2017). SVM was used to predict party affiliations from speeches or debate transcripts in the European Parliament (Høyland et al., 2014). From the observers of the US Supreme Court's voting bodies, it has been predicted whether the vote can be known based on the statements, discussions, and questions posed by voters (Budhwar et al., 2018). Other works present sentiment analysis for predicting the political arena through political speeches and debates (Abercrombie & Batista-Navarro, 2020). This study was conducted in the Republic of North Macedonia, specifically in the Assembly of North Macedonia (Assembly of the Republic of North Macedonia, 2026). The Probability of Law Approval (PLA) is an important aspect of parliamentary decision-making, as the approval or rejection of a law is influenced by numerous procedural, political, and institutional factors (Sahin, 2020; Naidu, Zuva, & Sibanda, 2023).

This study will answer the 3 research questions presented below:

RQ1: How effective are machine learning models for PLA in e-Parliament?

*Corresponding author: Safije Sadiki Shaini, ss29749@seeu.edu.mk

RQ2: Which procedural, political, and institutional variables contribute the most to the Probability of Law Approval (PLA) in an e-parliamentary system?

RQ3: Which ML model has the best performance for PLA prediction?

## 2. METHODOLOGY

### *2.1. Dataset Preparation and Cleaning*

This study was conducted in the Legislative Assembly of the Republic of North Macedonia. The procedural and legislative data used in this paper were obtained from public sources covering the period from 2008 to 2026. A dataset was created from scratch, containing 11,207 initial data points. After cleaning and processing, 8,704 data points from the dataset were worked with. Some of the column data are presented below: material id, title mk, registration no, registration date, status, type, procedure type, proposer, termination date, termination status, n FR sittings, n SR sittings,in TR sittings, first sitting date, last sitting date, n amendments FR, n amendments SR, in committees, etc. (Assembly of the Republic of North Macedonia, 2026). After cleaning and preprocessing, the final dataset contained 5,421 approved proposals and 3,283 non-approved proposals. These preprocessing steps were applied to improve data quality and reduce the risk of data leakage during predictive modeling (Hassler et al., 2019; Pal, 2022; Bernett et al., 2024).

### *2.2 Construction of the Target Variable*

The Probability of Law Approval (PLA) is formulated as a binary classification variable, since a law is either approved (1) or not approved (0). The data without valid information that could affect PLA results were removed and cleaned from the dataset. The target variable was constructed using parliamentary status information and final voting outcomes. This enabled the transformation of parliamentary approval into a classification problem suitable for supervised Machine Learning models.

### *2.3 Machine Learning Model Design*

In this study, several machine learning models were evaluated to identify the best model for PLA prediction. The evaluated models included:

- Logistic Regression
- Random Forest Classifier
- Gradient Boosting Classifier
- Support Vector Machine (SVM)
- XGBoost Classifier

The predictive variables included procedural features, amendment activity indicators, government and coalition variables, parliamentary workload measures, crisis-period indicators, and fast-track procedure indicators.

### *2.4 Model Evaluation and Validation*

To measure the performance of the models, the following metrics were used:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

An 80/20 train-test split was applied to preserve class-distribution consistency between the training and testing datasets (Indranie et al., 2025; Prusty, Patnaik, & Dash, 2022; Li et al., 2020). Because the dataset showed class imbalance between approved and non-approved law proposals, F1-score and ROC-AUC were considered very important for model evaluation (Gaudreault, Branco, & Gama, 2021). All models were evaluated using the same preprocessing and evaluation pipeline to give us a performance comparison across different machine learning approaches. To address class imbalance, SMOTETomek was applied only to the training data as part of an imbalanced-learn pipeline (Wang et al., 2019). In addition, class-weight balancing was used for Logistic Regression, Random Forest, and SVM.

### *2.5 Feature Engineering and Predictive Variables*

The purpose of feature engineering is to select features that are procedurally and parliamentary relevant and to avoid data leakage. To construct the features, key columns from the dataset were selected, including: type, procedure type, proposer, n amendments total,in committees, government, has majority, coalition size, is technical, is crisis, alb partner, is fast track,has eu flag, monthly workload, is holiday season. The feature selection and engineering strategy focused on selecting procedurally meaningful attributes while preventing data leakage (Popov, 2023). To predict the probability of law approval, we also included other variables relevant to whether laws will be voted on. During the years 2008-2026, there were many political dynamics, such as: the status of the government majority, the size of the coalition, technical periods of government, periods of crisis and partners of the Albanian coalition. These variables directly affect the PLA and are included in the dataset.

# 3. EXPERIMENTAL RESULTS

## *3.1 Model Comparison and Evaluation*

From Table 1, Random Forest achieved the best overall performance for PLA prediction, with an Accuracy of 0.847, an F1-score of 0.876, and a ROC-AUC value of 0.910. XGBoost and Gradient Boosting also showed good classification performance, whereas Logistic Regression performed worse than the ensemble-based models. The results reported in Table 1 were obtained

| Model | Accuracy | Precision | Recall | F1 | ROC-AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.773 | 0.818 | 0.817 | 0.818 | 0.827 |
| Random Forest | 0.847 | 0.881 | 0.872 | 0.876 | 0.910 |
| Gradient Boosting | 0.820 | 0.867 | 0.839 | 0.853 | 0.889 |
| SVM | 0.805 | 0.858 | 0.822 | 0.840 | 0.855 |
| XGBoost | 0.828 | 0.870 | 0.851 | 0.861 | 0.899 |

**Table 1**: *Model Performance Results for PLA Prediction*

using an 80/20 train-test split, where 80% of the data were used for training and 20% for testing. Stratified 5-Fold Cross-Validation was additionally used to evaluate model stability and generalization performance across multiple training iterations (Wang et al., 2019). Cross-validation results showed good performance across the evaluated folds, particularly for the ensemble-based models (Popov, 2023; Dong et al., 2020).

## *3.2 Confusion Matrix Analysis*

From the figure, we observe that the laws approved and correctly predicted by the model are True Positives (945), indicating that the model has a strong ability to identify laws that pass and are voted on. For laws that were not approved, the True Negatives (529) were correctly predicted (Polikar, 2012). There are False Positives (128), meaning laws that were not approved but were predicted by the model as approved. Finally, there are 139 approved laws that were predicted as not approved

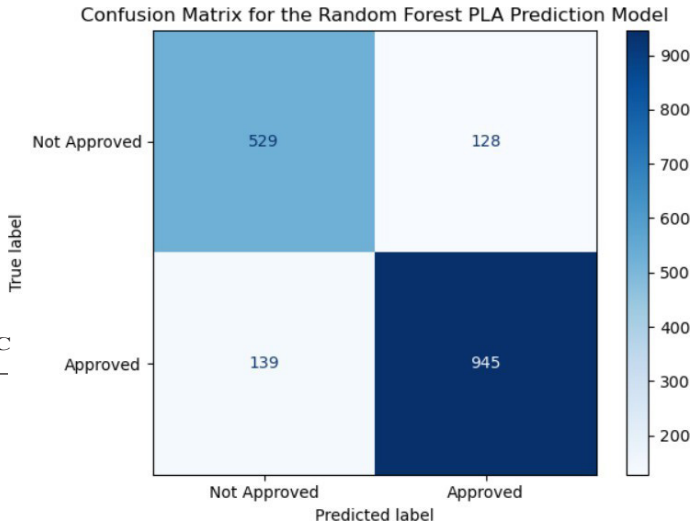



**Figure 1** :*Confusion Matrix for the Random Forest PLA Prediction Model*

## *3.3 Feature Importance Analysis*

In Figure 2 we present the feature importance analysis for the Random For- est model, where variables related to the proposer of laws, particularly proposals submitted by Members of Parliament (proposer MP), Working Bodies (pro- poser Working Body), and the Government of the Republic of North Macedonia (proposer Government), together with procedural, political, and institutional variables, had an important impact on PLA prediction. Political and coalition-related indicators also contributed to the model, confirming that legislative approval is shaped by the interaction among procedural, institutional, and political factors rather than by a single dominant variable (Konrath & Rathmair, 2025).

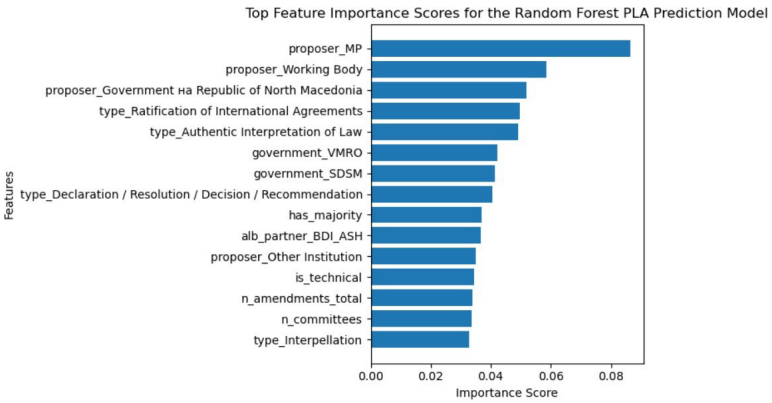



**Figure 2** :*Top Feature Importance Scores for the Random Forest PLA Prediction Model*

## 4. CONCLUSION

From this study, we conclude that Machine Learning models can predict the Probability of Law Approval (PLA) in e-Parliament systems, providing a direct answer to RQ1. Procedural, political, institutional, and proposer-related variables significantly influenced PLA prediction, answering RQ2. Among the evaluated models, Random Forest achieved the best overall performance for PLA prediction. The results showed that ensemble-based models achieved better classification performance than linear models, thereby answering RQ3. Overall, the findings show that ML techniques can support parliamentary decision-making and contribute to the development of more advanced e-Parliament systems.

## 5. LIMITATIONS

Despite the good classification performance, this study has several limitations. This dataset does not include the texts of laws, amendments, and parliamentary debates. Also, the dataset imbalance affected the performance of the models used in this paper.

## 6. FUTURE WORK

In the future, the texts of laws and amendments may also be included, in which Natural Language Processing (NLP) methods can be applied to improve the predictive performance of the models within an e-Parliament System.

## 7. REFERENCES

Abercrombie, G., & Batista-Navarro, R. T. (2020). ParlVote: A corpus for sentiment analysis of political debates. In Proceedings of the 12th Language Resources and Evaluation Conference (pp. 5073–5078).

Assembly of the Republic of North Macedonia. (2026). Official parliamentary website. https://www.sobranie.mk

Arslan, S. (2025). The impact of parliamentary oversight mechanisms on foreign policy: A legal and democratic framework. International Journal of Educational Research, 8(11), 01–20.

Bernett, J., Blumenthal, D. B., Grimm, D. G., Haselbeck, F., Joeres, R., Kalinina, O. V., & List, M. (2024). Guiding questions to avoid data leakage in biological machine learning applications.

Budhwar, A., Kuboi, T., Dekhtyar, A., & Khosmood, F. (2018). Predicting the vote using legislative speech. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age (pp. 1–10).

Cavalieri, A., Ducange, P., Fabi, S., Russo, F., & Tonellotto, N. (2022). An intelligent system for the categorization of question time official documents of the Italian Chamber of Deputies. Journal of Information Technology & Politics, 1–22.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14(2), 241–258.

Fitsilis, F., Koryzis, D., & Schefbeck, G. (2022). Legal informatics tools for evidence-based policy creation in parliaments. International Journal of Parliamentary Studies, 2(1), 5–29.

Gaudreault, J.-G., Branco, P., & Gama, J. (2021). An analysis of performance metrics for imbalanced classification. In International Conference on Discovery Science (pp. 67–77). Springer.

Goldblatt, D., & O'Neil, T. (2012). How a bill becomes a law—Predicting votes from legislation text. Technical report, Stanford University.

Hassler, A. P., Menasalvas, E., García-García, F. J., Rodríguez-Mañas, L., & Holzinger, A. (2019). Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome. BMC Medical Informatics and Decision Making, 19(1), 33.

Høyland, B., Godbout, J.-F., Lapponi, E., & Velldal, E. (2014). Predicting party affiliations from European Parliament debates. In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science (pp. 56–60).

Indranie, C. A., Ramadhan, A., Raharjo, J., Ikhsan, R. R. N., Rohadi, M. F., & Anandaputra, D. I. (2025). Impact of train-test splitting ratios on machine learning models for daily electricity load time series forecasting. Presented in 2025 2nd International Conference on Information System and Information Technology (ICISIT) (pp. 1–6). IEEE.

Li, S., Chen, Z., Liu, Q., Shi, W., & Li, K. (2020). Modeling and analysis of performance degradation data for reliability assessment: A review. IEEE Access, 8, 74648–74678.

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. In Computer Science Online Conference (pp. 15–25). Springer.

Nay, J. J. (2017). Predicting and understanding law-making with word vectors and an ensemble model. PLOS ONE, 12(5), e0176999.

Pal, A. (2022). DeepParliament: A legal domain benchmark & dataset for parliament bills prediction. In Proceedings of the Workshop on Unimodal and Multimodal Induction of Linguistic Structures (UM-IoS) (pp. 73–81).

Polikar, R. (2012). Ensemble learning. In Ensemble Machine Learning (pp. 1–34). Springer.

Popov, A. (2023). Feature engineering methods. In Advanced Methods in Biomedical Signal Processing and Analysis (pp. 1–29). Elsevier.

Prusty, S., Patnaik, S., & Dash, S. K. (2022). SKCV: Stratified k-fold cross-validation on ML classifiers for predicting cervical cancer. Frontiers in Nanotechnology, 4, 972421.

Reis, J., Santo, P. E., & Melão, N. (2019). Impacts of artificial intelligence on public administration: A systematic literature review. In 2019, 14th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1–7).

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308.

Sulea, O. M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., & Van Genabith, J. (2017). Exploring the use of text classification in the legal domain. arXiv preprint arXiv:1710.09306.

Wang, Z., Wu, C., Zheng, K., Niu, X., & Wang, X. (2019). SMOTETomek-based resampling for personality recognition. IEEE Access, 7, 129678–129689. https://doi.org/10.1109/ACCESS.2019.2940061

Konrath, C., & Rathmair, A. (2025). Parliaments in the digital age—Building blocks for a theoretical framework. In Interdisciplinary Science and Research Conference on Digital Humanism (pp. 280–294). Springer.

crj.cit.edu.al

CIT
CANADIAN INSTITUTE OF TECHNOLOGY

# CIT REVIEW JOURNAL

# C | R | J

PUBLISHER
**Canadian Institute of Technology**
Tirana, Albania