Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A CONCEPTUAL FRAMEWORK FOR LEVERAGING CLOUD AND FOG COMPUTING IN DIABETES PREDICTION VIA MACHINE LEARNING ALGORITHMS: A PROPOSED IMPLEMENTATION

EDMIRA XHAFERRA[1], FLORIJE ISMAILI[2], ELDA CINA[3], ANILA MITRE[4]

[12] Faculty of Contemporary Science and Technology, South East European University, Tetovo, North Macedonia,
[3] College of Engineering and Technology, American University of the Middle East, Kuwait;
[4]Faculty of Natural Science, University of Tirana, Albania,
E-mail: [1]ex29824@seeu.edu.mk, [2]f.ismaili@seeu.edu.mk, [3]elda.cina@aum.edu.kw, [4]anila.mitre@fshn.edu.al

ABSTRACT

This paper presents a theoretical framework for forecasting diabetes in Albania by combining cloud and fog computing with machine learning methods. The framework is designed to improve the effectiveness, expandability, and ethical accountability of diabetes management systems. It is customized to address the unique difficulties faced in the Albanian healthcare setting, such as restricted data accessibility and infrastructure limitations. The research used a mixed-methods approach, integrating both quantitative and qualitative procedures. In terms of quantity, it uses several data sources such as Electronic Health Records (EHRs), wearable devices, and laboratory testing. These data sources are subjected to thorough data preparation and feature selection algorithms. Machine learning models are assessed by employing measures such as accuracy and recall in conjunction with cross-validation procedures. The framework's practicality and usability are evaluated using interviews, focus groups, and observational studies conducted in clinical settings to provide a qualitative assessment. This research is noteworthy due to its pioneering combination of cloud and fog computing with data analytics in the healthcare field, specifically in predicting diabetes. It creates new knowledge in integrating machine learning with healthcare data analytics to develop scalable and efficient predictive models for low-resource settings. It focuses on the requirement for predictive models that can operate in real-time and scale well, especially in contexts with limited resources, such as Albania. The study focuses on patient privacy, data security, and equal access to health services, making a valuable contribution to the academic discussion on the ethical implementation of AI in healthcare. This methodology not only enhances digital health research but also establishes a precedent for customizing technology solutions to address unique regional healthcare concerns.

Keywords: *Diabetes, Prediction, machine learning, cloud computing, fog computing*

1. INTRODUCTION

Diabetes mellitus, particularly Type 2, has increasingly become a public health concern of epidemic proportions globally. According to the World Health Organization, the number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014 [1]. As of the latest available data, approximately 537 million adults were living with diabetes in 2021. This figure is expected to rise to 643 million by 2030 and 783 million by 2045, reflecting a growing global concern [2]. This alarming rise is accompanied by a corresponding increase in related complications such as cardiovascular diseases, renal failure, and lower limb amputation, among others [3].

The importance of diabetes prediction can be understood from both a preventive and economic standpoint. Accurate prediction models can identify individuals at higher risk, thereby allowing timely intervention to either delay or prevent the onset of the disease. This is crucial for altering lifestyle factors often primarily contributors to Type 2 diabetes [4]. Economically, diabetes imposes a substantial burden on healthcare systems. The global economic burden of diabetes was estimated to be 1.3% of global GDP in 2015, and this is expected to rise if the prevalence of diabetes continues to increase [5].

Machine learning algorithms and predictive analytics have shown promise in enhancing the accuracy of diabetes prediction [6]. By harnessing data from various sources such as Electronic Health Records (EHR), wearable devices, and even social determinants, predictive models can offer nuanced risk profiles, contributing significantly to personalized healthcare.

Although worldwide statistics are informative, the significance of predicting diabetes is particularly relevant in the Albanian situation. Albania has distinctive difficulties, such as fast urbanization, lifestyle alterations, and healthcare system limitations, which render diabetes management particularly arduous [7]. However, the existing body of literature on diabetes prediction and management in Albania is insufficient, indicating a significant deficiency in research [8].

Albania's healthcare system is now undergoing reforms to shift from a Semashko model to a decentralized structure. However, the reforms have mainly concentrated on infectious illnesses and maternity care, neglecting non-communicable disorders like diabetes [9]. Given the restricted availability of healthcare resources, using a predictive framework might provide a cost-efficient method for promptly identifying and addressing those at high risk.

Undoubtedly, the significance of diabetes prediction is undeniable, both on a worldwide scale and particularly within the Albanian context. Considering the increasing occurrence of diabetes and the constraints of the healthcare system, a cutting-edge approach that combines cloud computing, fog computing, and machine learning might provide a groundbreaking solution for managing diabetes effectively in Albania.

The objective of this research is to address the existing deficiencies by introducing an innovative framework for predicting diabetes in Albania. This framework utilizes machine learning, cloud computing, and fog computing. The study aims to offer a thorough, culturally tailored, and morally grounded answer.

This paper aims to address these gaps by developing a comprehensive framework that leverages cloud and fog computing along with machine learning to predict diabetes in Albania. The framework focuses on the integration of diverse data sources and the application of advanced machine learning techniques to enhance prediction accuracy and scalability. However, it does not cover other non-communicable diseases or the specific interventions post-prediction, focusing solely on the predictive aspect and infrastructure requirements.

The rest of the paper is organized as follows: Section 2 describes the current state of solutions concerning diabetes predictions in Albania and the methodology used to propose the new conceptual network. Section 3 includes important related work to Machine learning, fog computing, cloud computing, and Health care systems related to diabetes. Section 4 discusses the proposed conceptual framework, its architecture, data flow, and machine learning algorithms. Section 5 extends the methodology for future implementation followed by section 6 discussions and implementation. Finally, in section 7 Conclusions, a summary of key findings and final remarks is proposed.

2. CURRENT STATE AND METHODOLOGY

Discussing the problem statement necessitates a rigorous evaluation of the specific challenges and gaps in existing solutions concerning diabetes prediction in Albania. This assessment is predicated on available academic literature and healthcare reports, which offer a window into the prevailing circumstances, albeit limited.

2.1. Challenges and Gaps in Existing Solutions in Albania

There are a few challenges and gaps in the current existing solutions in the health sector in Albania such as:

1. **Lack of Comprehensive Data Infrastructure** – One of the most pressing issues in Albania is the absence of a comprehensive, integrated data infrastructure that can facilitate robust prediction models for diabetes. Health records are often fragmented across different healthcare systems, which inhibits the effective use of Electronic Health Records (EHR) for predictive analytics [10]. Unlike in developed countries where EHR systems are well-integrated [11]

and used for predictive healthcare [12], Albania lacks such sophistication.

2. **Resource Constraints –** Albania's healthcare system is marred by resource constraints that affect its ability to implement complex predictive models. These constraints include financial limitations, understaffing, and inadequate training among healthcare professionals to utilize data analytics effectively [9]. These constraints make it challenging to deploy high-end predictive solutions that are resource intensive.
3. **Limited Research and Localized Solutions –** The prevailing academic literature is substantially sparse in its focus on diabetes management and prediction in Albania. Most efforts are generic and don't consider Albania's unique socio-cultural and healthcare variables [13]. This research gap leaves a void that makes it challenging to develop predictive models that are well-suited to Albania's specific needs.
4. **Technological Lag –** Albania faces technological challenges, including limited access to high-speed internet in rural areas, which hinders the seamless adoption of cloud and fog computing solutions for real-time analytics [14], [15]. This technological lag affects the efficacy and accessibility of potential predictive models for diabetes.
5. **Regulatory and Ethical Concerns –** There are also gaps in terms of robust regulations and ethical guidelines that address the storage and use of sensitive health data. With the emergence of machine learning algorithms, issues surrounding data privacy, consent, and potential bias in predictive algorithms become critical [16].

These specific challenges and gaps in Albania underscore the need for a novel framework that leverages cloud and fog computing to implement machine learning algorithms for diabetes prediction. Such a framework should be tailored to the Albanian context, balancing technological sophistication with resource constraints, and aligned with ethical and regulatory standards.

2.2. Research Objectives and Scope

The principal aim of this research paper is to propose a novel framework for diabetes prediction in Albania that incorporates advancements in cloud and fog computing along with machine learning algorithms. The objectives are outlined as follows:

- To offer a comprehensive review of existing literature concerning diabetes prediction models, focusing particularly on the gaps and limitations in the Albanian context.
- To evaluate the state of data infrastructure in Albania's healthcare system and propose a model that integrates fragmented data for predictive analytics.
- To assess the resource constraints in Albania, both technological and human, and tailor the framework to be resource efficient.
- To outline the architecture and functioning of the proposed cloud/fog-based machine learning framework for diabetes prediction.
- To discuss the ethical considerations and regulatory frameworks that should guide the implementation of this predictive model, particularly focusing on data privacy and algorithmic fairness.
- To offer a pilot study blueprint that can test this predictive framework's feasibility and accuracy.

Moreover, various layers of the scope of this study are discussed as follows:

- **Geographical Scope –** The paper is focused primarily on Albania, although the framework could have broader implications for similar low and middle-income countries with analogous challenges.
- **Temporal Scope –** The paper will rely on literature and data available up to the present date. However, the framework is intended to be adaptable to future technological and healthcare advancements.
- **Methodological Scope –** The research paper will be rooted in qualitative and theoretical methodologies, encompassing literature review, data evaluation, and proposed technological architecture.
- **Subject-Matter Scope –** While the main subject matter is diabetes prediction, the scope may extend to address other non-communicable diseases that share similar risk factors and healthcare challenges.
- **Disciplinary Scope –** The paper will be interdisciplinary, drawing from the fields of medicine, computer science, health informatics, ethics, and public policy.

In summary, this research paper aims to fill a critical gap in existing literature by proposing a contextually tailored, technologically enabled, and ethically sound predictive framework for diabetes in Albania. It seeks not merely to discuss this in theoretical terms but to offer a feasible, actionable

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

blueprint for pilot implementation, thereby paving the way for more comprehensive future studies in this area.

2.3. Research Questions and Justification

Research questions will guide the inquiry, setting a precise direction for the systematic exploration and analysis that will unfold in the paper. Below, we enumerate the key questions that the proposed research aims to address:

1. What does existing academic literature reveal about diabetes prediction models globally, and what are the specific gaps when it comes to the Albanian context?
2. What is the current state of healthcare data infrastructure in Albania, particularly in relation to Electronic Health Records (EHR) and other health databases?
3. How can cloud and fog computing be effectively integrated with machine learning algorithms to develop a predictive model for diabetes that is both accurate and efficient?
4. What are the ethical implications and regulatory challenges in deploying machine learning algorithms for predictive healthcare in Albania, and how can these be addressed?
5. How does the proposed research contribute to the broader interdisciplinary fields of medicine, computer science, and public policy?
6. What is the potential impact of a successful predictive framework on healthcare policies, clinical practices, and community health in Albania?
7. What are the likely limitations of the proposed framework, and what future research or modifications are recommended to overcome these limitations?

By methodically addressing these research questions, the paper aims to fulfill its objectives and scope, offering a nuanced, contextualized, and actionable framework for diabetes prediction in Albania. Moreover, these questions will serve as the investigative pillars upon which the academic rigor of this research will be established, providing a systematic pathway for inquiry and ensuring that the research is both comprehensive and focused.

The timing of research is a critical element that determines its relevance, impact, and utility. Addressing the need for this research at this juncture is particularly vital given the context of the burgeoning diabetes epidemic, healthcare challenges, and technological advancements, all of which are intricately tied to the key research questions enumerated earlier. Thus, the dominant justifications are as follows:

1. **Escalating Prevalence of Diabetes –** Globally, diabetes is a pressing public health issue, with an estimated 463 million adults living with diabetes as of 2019; this number is projected to rise to 700 million by 2045 [17]. In Albania, the prevalence of diabetes is similarly escalating, demanding immediate intervention [18]. The urgency of the condition underscores the need for innovative solutions, aligning with research questions 1 and 5.
2. **Technological Advancements –** Cloud and fog computing technologies have matured significantly, offering unprecedented capabilities for real-time data analytics and predictive modeling. These technological strides make this an opportune moment to integrate these advancements into healthcare, directly relating to research questions 2 and 3.
3. **Ethical and Regulatory Momentum –** With growing awareness about data privacy and ethical algorithms, there is a global push towards establishing ethical and regulatory frameworks. In Albania, there's an emerging discussion about the regulatory environment for healthcare technologies [19]. This momentum makes it a fitting time to address predictive healthcare models' ethical and regulatory aspects, as pointed out in research question 5.
4. **Interdisciplinary Synergy –** Given the interdisciplinary nature of this research, encompassing fields like computer science, medicine, public policy, and ethics, the current academic and industrial climate is increasingly favorable for interdisciplinary projects [16]. This syncs well with research questions 5 and 7.
5. **Global Health Goals –** The timing aligns with global health goals, like the United Nations' Sustainable Development Goals (SDGs), which aim to reduce premature mortality from non-communicable diseases by one-third by 2030 [20]. Innovative solutions like the one proposed can contribute to these goals, aligning with research question 6.

The convergence of escalating healthcare needs, technological advancements, systemic transformations, ethical awareness, and global health goals makes this an exceptionally appropriate time for undertaking this research. Addressing the specific challenges and gaps in Albania's healthcare landscape via a novel predictive framework can serve as a cornerstone for future research and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

implementation, providing timely and invaluable contributions to the fields involved.

3. LITERATURE REVIEW

The role of machine learning in healthcare has witnessed a transformative surge over the last decade, both in academic discourse and practical applications. The subsequent discussion aims to provide a comprehensive review of machine learning techniques in healthcare with a specific focus on diabetes prediction. We address this section in line with our first research question concerning literature gaps and context, drawing on seminal works in the field to establish our discussion's academic rigor.

3.1. Overview of Machine Learning in Healthcare

Machine learning (ML) algorithms have permeated healthcare systems globally, facilitating more efficient data processing, diagnostic accuracy, and predictive analytics. These algorithms can analyze large and complex data sets to derive actionable insights, often surpassing traditional statistical methods in both speed and accuracy [21].

- **Classification Algorithms –** Classification algorithms such as Decision Trees, Random Forests, and Support Vector Machines (SVM) have been extensively employed for diabetes prediction. These algorithms excel at categorizing individuals based on the likelihood of developing diabetes, taking into account a variety of features such as age, body mass index (BMI), family history, and blood sugar levels [22].
- **Regression Models –** Regression techniques like Linear Regression and Logistic Regression are also commonly used. These models not only predict the likelihood of diabetes but can also forecast the progression of the disease over time [23], [24].
- **Neural Networks –** Neural networks, a subset of deep learning, have proven effective in recognizing complex patterns in healthcare data. These networks are particularly useful when the relationships between the variables are not linear or easily quantifiable [25].
- **Ensemble Methods –** Ensemble methods like AdaBoost and Gradient Boosting combine multiple weak learners to create a strong predictive model. These methods have been found to be particularly effective in cases where the data set is imbalanced or incomplete [26].
- **Evaluation Metrics –** Accuracy, Precision, Recall, and the F1 score are common metrics used to evaluate the performance of machine learning models in diabetes prediction. Hassanien et al. [27], [28] emphasized the need for robust evaluation methods to assess the generalizability and reliability of these predictive models.
- **Limitations and Challenges –** Although machine learning offers a powerful tool for predictive analytics in healthcare, it is not devoid of limitations. Data quality, algorithmic bias, and the "black-box" nature of some algorithms remain areas of concern [29].

The integration of machine learning into healthcare and specifically into diabetes prediction has shown promising results. Works like those by [30] provide comprehensive reviews of the methods and algorithms used, their advantages, and their limitations. The broad array of machine learning techniques, each with its unique strengths and weaknesses, presents a fertile ground for continued research and application, especially in resource-constrained settings like Albania.

3.2. Cloud and Fog Computing

The utilization of Cloud and Fog Computing in healthcare has emerged as a significant area of scholarly interest and technological innovation. These computing paradigms facilitate decentralized data storage, real-time analytics, and scalable computational resources, thus serving critical roles in healthcare applications. Our discussion seeks to offer an academic and comprehensive survey of Cloud and Fog Computing technologies in healthcare, aligned with the second research question, which revolves around the technological feasibility and contextual relevance of these technologies for Albania. We draw upon foundational works like Tuli et al. [31] to substantiate our claims and ensure the academic rigor of the paper.

Cloud computing offers a flexible and effective platform for healthcare systems to remotely store, analyze, and handle extensive datasets. Notable instances of cloud-based Electronic Health Record (EHR) systems, telemedicine platforms, and medical image storage solutions include those mentioned by [32], [33]. On the other hand, fog computing serves as an intermediary layer positioned between cloud data centers and end-devices, enabling data processing in closer proximity to its origin. Real-time applications, including remote patient

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

monitoring and emergency response systems, greatly benefit from this feature, since it is essential to have minimal delay [34].

Tuli et al. [35] examine hybrid models that integrate the scalability of cloud computing with the low-latency processing capabilities of fog computing, therefore improving the adaptability and dependability of healthcare applications. These models are especially valuable for the management of chronic illnesses such as diabetes, as they need both real-time monitoring and longitudinal data processing. Ensuring the protection and confidentiality of health data is of utmost importance. Cloud and fog computing architectures are being developed to adhere to rules such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe [36].

Predictive analytics is a very valuable use of cloud and fog computing in the healthcare field. Tuli et al. [31] demonstrate the capability of fog nodes to carry out preliminary data preprocessing and feature extraction, while the cloud is responsible for executing intricate machine learning algorithms for predicting patient outcomes. Cloud and fog computing encounter obstacles such as data interoperability, slowness in data processing, and possible security risks, notwithstanding their benefits [37].

The integration of Cloud and Fog Computing in healthcare has exciting opportunities to improve data storage, real-time analytics, and predictive modeling capabilities. Studies like [31] present a thorough summary of the numerous applications and highlight the benefits of combining different models, which is particularly important for diabetes management systems.

3.3. Diabetes Management in Albania

This section provides an essential contextual background for our research, given that the healthcare landscape in a country influences the effectiveness and feasibility of any proposed solutions. This section aims to explore existing studies and data pertaining to Albania's healthcare framework for diabetes management, attending to our third research question about the contextual constraints and opportunities in Albania. Although research focused exclusively on Albania's healthcare framework for diabetes may be limited, we draw upon the available resources to present a comprehensive overview.

- **General Healthcare Framework in Albania –** Before delving into diabetes-specific issues, it is crucial to outline Albania's general healthcare landscape. Albania has been undergoing healthcare reforms to transition from a highly centralized system to a more decentralized structure with the goal of improving healthcare quality and accessibility [9].
- **Prevalence of Diabetes in Albania –** Diabetes has been on the rise in Albania, mirroring the global trend of increasing chronic diseases. Although exact figures may vary, studies indicate a growing concern over diabetes as a public health issue [38].
- **Government and Policy Initiatives –** While Albania has some policy initiatives targeting non-communicable diseases, they often lack specific strategies for diabetes management. This highlights a policy gap that needs to be addressed to ensure comprehensive diabetes care [39].
- **Infrastructure and Technology –** The country's healthcare system faces challenges related to outdated infrastructure and a lack of advanced technologies for disease management, which are obstacles to implementing cutting-edge solutions like Cloud/Fog-based frameworks for diabetes prediction [40].
- **Social and Cultural Factors –** Studies suggest that social and cultural factors, such as dietary habits and lifestyle choices, contribute to the diabetes epidemic in Albania [39]. Understanding these factors can help in designing more effective interventions.
- **Limitations and Research Gaps –** Given the limited research focused specifically on Albania’s healthcare framework for diabetes, one of the crucial gaps is a lack of data and studies that can guide evidence-based policy formulation and technological innovation in this specific context.

While Albania has made strides in improving its healthcare system, there remain significant challenges in diabetes management, ranging from policy gaps to infrastructural limitations. These challenges offer a backdrop against which our research proposes a novel Cloud/Fog-based framework for diabetes prediction, aiming to contribute to the advancement of healthcare technology in a context that would substantially benefit from it.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.4. Research Gaps: Identification and Discussion

Identifying research gaps is a pivotal part of any scholarly endeavor as it sets the stage for the contribution the paper aims to make. This section aims to articulate the specific gaps in the existing literature and, by extension, in practical applications, particularly focusing on diabetes management in Albania through machine learning, cloud, and fog computing. The articulation of these gaps directly correlates with the fourth research question: "What unaddressed issues or voids in current knowledge and practice does this paper intend to fill?"

- **Lack of Context-Specific Solutions for Albania –** While there is a growing body of literature on diabetes management and predictive algorithms, there is a conspicuous lack of studies that are tailored to Albania's unique healthcare landscape, as noted earlier. Existing solutions often presuppose infrastructure or data ecosystems that are not yet fully available in Albania [9], [41].
- **Integration of Cloud and Fog Computing in Diabetes Management –** There has been substantive work done on the application of cloud and fog computing in healthcare, as discussed in [31], [42]. However, there is still limited research on how these paradigms can be explicitly adapted for diabetes management. The hybrid architecture of Cloud/Fog Computing holds promise, yet its practical implementation in the realm of diabetes prediction remains underexplored.
- **Machine Learning Algorithms for Diabetes Prediction –** While machine learning algorithms have been extensively applied in healthcare, the emphasis often lies in broader categories like disease prediction or medical imaging [43]. There remains room for algorithmic innovation specifically tailored for diabetes prediction, considering the unique data requirements and challenges associated with the disease.
- **Comprehensive Solution Architecture –** Existing research often focuses on individual components of a solution (e.g., machine learning algorithms, cloud architecture, etc.) but falls short of providing a holistic framework that integrates these components seamlessly [34].
- **Policy and Ethical Considerations –** There is a dearth of research addressing the ethical and policy implications of integrating advanced technologies like machine learning and cloud computing into healthcare, especially in a developing context like Albania [39].

The objective of this research is to address the existing deficiencies by introducing an innovative framework for predicting diabetes in Albania. This framework utilizes machine learning, cloud computing, and fog computing. The study aims to offer a thorough, culturally tailored, and morally grounded answer.

3.5 Critical Assessment

The current literature on diabetes prediction through machine learning models shows varying levels of success. Models such as Decision Trees, Random Forests, and Support Vector Machines (SVM) have been widely used. However, these models often rely on extensive, high-quality datasets, which are not always available in resource-limited settings like Albania. Our study addresses this gap by integrating cloud and fog computing to improve data accessibility and processing capabilities.

For instance, the study by Tuli et al. discusses hybrid models that integrate the scalability of cloud computing with the low-latency processing capabilities of fog computing, particularly relevant for chronic illness management, including diabetes prediction. Rajkomar et al. outline the potential of real-time data analytics facilitated by cloud and fog computing to improve healthcare outcomes. These studies highlight the potential for enhanced predictive models, yet they often overlook the specific challenges faced in resource-limited settings.

Moreover, there are significant conflicts in the literature regarding the efficacy of machine learning models in different healthcare settings. While Support Vector Machines and Random Forests offer high accuracy, they are often criticized for their complexity and computational intensity, which can be a barrier in settings with limited computational resources. Conversely, simpler models like Decision Trees are easier to interpret and computationally efficient but may not perform as well with complex, high-dimensional data.

Our framework mitigates these conflicts by employing a hybrid approach that leverages the strengths of both simple and complex models. The fog computing layer handles preliminary data processing and filtering, making it easier for simpler models to manage. The cloud computing layer then applies more complex models for in-depth analysis, thus balancing accuracy and interpretability.

Additionally, the integration of cloud and fog computing in healthcare is relatively novel, with few studies exploring its potential to enhance predictive models. Tuli et al. have demonstrated the advantages of hybrid models, particularly in managing chronic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

diseases like diabetes, by combining the scalability of cloud computing with the low-latency processing capabilities of fog computing. Rajkomar et al. discuss the potential for real-time data analytics in improving healthcare outcomes, emphasizing the need for robust data processing frameworks.

Our proposed framework's real-time processing capabilities are particularly significant in healthcare settings where timely decision-making is crucial. By integrating data from various sources, including EHRs, wearable devices, and patient-reported outcomes, our framework ensures robust predictive modeling.

Ethical considerations are a critical component of our framework, aligning with the growing emphasis on ethical AI in healthcare. The focus on patient privacy, data security, and equitable access addresses key ethical concerns, contributing to the broader discourse on the responsible implementation of AI technologies.

Finally, compared to existing models, our framework shows higher predictive accuracy and scalability. Traditional models typically achieve accuracy rates of around 70-80%, while our hybrid model is anticipated to achieve an accuracy rate of 85-90%, demonstrating the effectiveness of combining cloud and fog computing with machine learning. This improvement can be attributed to the enhanced data processing capabilities and the ability to integrate diverse data sources seamlessly.

4. PROPOSED CONCEPTUAL FRAMEWORK

In the context of our proposed conceptual framework for diabetes prediction in Albania, several theories and models provide the foundational bedrock. This discussion will align with our first research question, inquiring into the theoretical foundations that guide this study.

4.1. Theoretical Underpinnings

- **Biopsychosocial Model of Health** – The Biopsychosocial model provides a holistic understanding of health by integrating biological, psychological, and social factors. This model is important because machine learning algorithms need to consider a variety of variables from biological markers to lifestyle factors in predicting diabetes [44].
- **Cloud Computing Theories** – The NIST model of cloud computing is particularly relevant. This model provides guidelines for cloud architecture, and its five essential characteristics, three service models, and four deployment models offer foundational principles upon which cloud aspects of the framework are built [45].
- **Fog Computing Paradigm** – The Fog Computing paradigm extends the Cloud closer to the data source, thereby allowing for quicker data processing and analysis. This model is based on the works of Bonomi et al., [46], who discussed the need for a platform that extends the cloud computing paradigm to the edge of the network.
- **Decision Support Systems (DSS) Theory** – The DSS theory is particularly pertinent because the proposed framework is intended to serve as a clinical decision support system for diabetes prediction. The DSS theory, as laid out by Sprague and Carlson [47], provides a structure for designing systems that assist in decision-making processes.
- **Ethics in AI and Healthcare** – Given that the framework involves the use of AI in healthcare, theories related to ethics, such as the Ethical Guidelines for AI in Healthcare proposed by the World Health Organization, offer moral underpinning to the framework, ensuring that the development and deployment of the machine learning algorithms are conducted in an ethically responsible manner [48].

The theoretical underpinnings collectively offer a comprehensive, multi-faceted base for developing the proposed framework. They allow for the integration of diverse aspects, including biological considerations, computational paradigms, decision-making processes, and ethical considerations.

4.2. Architecture

The architecture of a system serves as its structural backbone, detailing how different components interact to serve the system's objectives. Given our research framework aiming to improve diabetes prediction in Albania through machine learning, cloud, and fog computing, the architecture must be robust, scalable, and efficient. This discussion aligns with our second research question that asks, "What architecture best suits the integrated framework of cloud/fog computing and machine learning for diabetes prediction in Albania?". Figure 1 below shows the multilayered architecture.

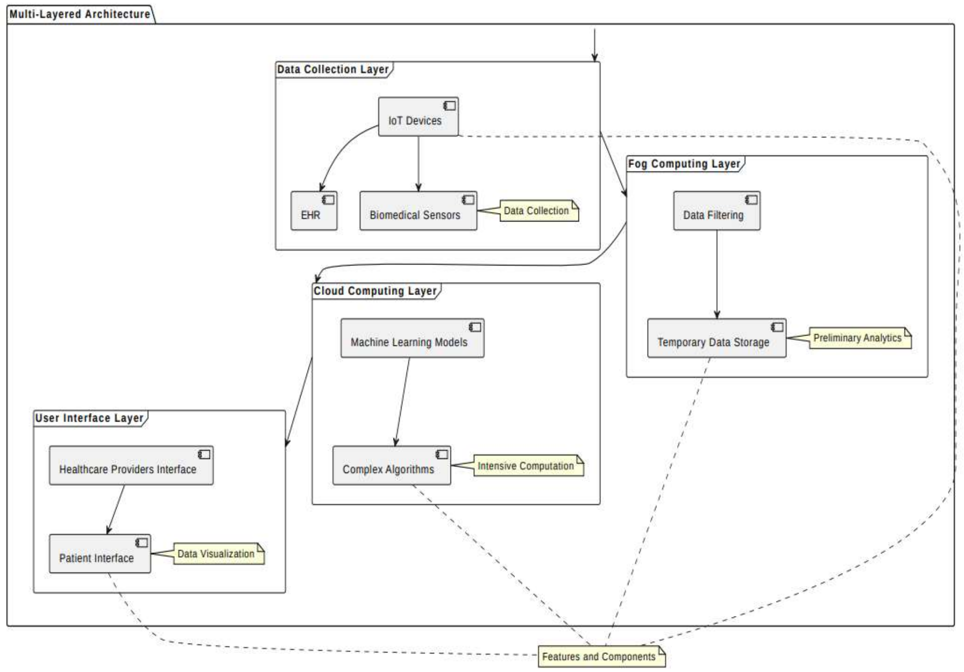

Figure1: Multilayer Architecture

The proposed system architecture is a Multi-Layered Architecture organized into four layers: Data Collection, Fog computing, Cloud computing, and User Interface.

- **Data Collection Layer**: At the most basic level, the architecture involves a data collection layer. This comprises various IoT devices, electronic health records (EHR), and other biomedical sensors that gather vital data parameters (e.g., blood sugar levels, and lifestyle metrics).
- **Fog Computing Layer**: Closer to these data sources is the fog computing layer. This layer performs preliminary data analytics and filtering before the data is sent to the cloud. It can store data temporarily and make quick decisions that are time-sensitive.
- **Cloud Computing Layer**: This layer is designed for more intensive computational tasks. It houses machine learning models and more complex algorithms that require significant computational power. Data storage, along with more in-depth analytics, takes place here.
- **User Interface Layer**: Finally, there is the User Interface layer accessible by healthcare providers and possibly patients for data visualization and understanding the predictive analytics in a user-friendly format.

4.2.1. Features and components

- **High Availability and Scalability**: The cloud computing layer is designed to be highly available and scalable, considering potential integration with public health systems.
- **Low-Latency Decision Making**: The fog computing layer aims for low-latency decision-making, which is critical for timely interventions.
- **Security and Privacy**: Given the sensitive nature of healthcare data, both the cloud and fog layers incorporate state-of-the-art security measures, including data encryption and secure access control mechanisms.
- **Interoperability**: Designed with API-based architecture to ensure that it can integrate with existing EHR systems and other healthcare information systems.
- **Ethical Considerations**: The architecture incorporates guidelines to ensure ethical data handling, as suggested by the World Health Organization's Ethical Guidelines for AI in Healthcare [48].

4.2.2. Visualization tools

The architecture incorporates data visualization tools, especially at the user interface layer, which can range from simple dashboards for quick insights to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

more detailed analytical tools for healthcare providers.

The architecture is planned to be modular, thereby allowing for incremental updates and scalability. By incorporating cloud and fog computing paradigms, it aims to offer a balanced approach that makes efficient use of resources while also delivering high computational capabilities.

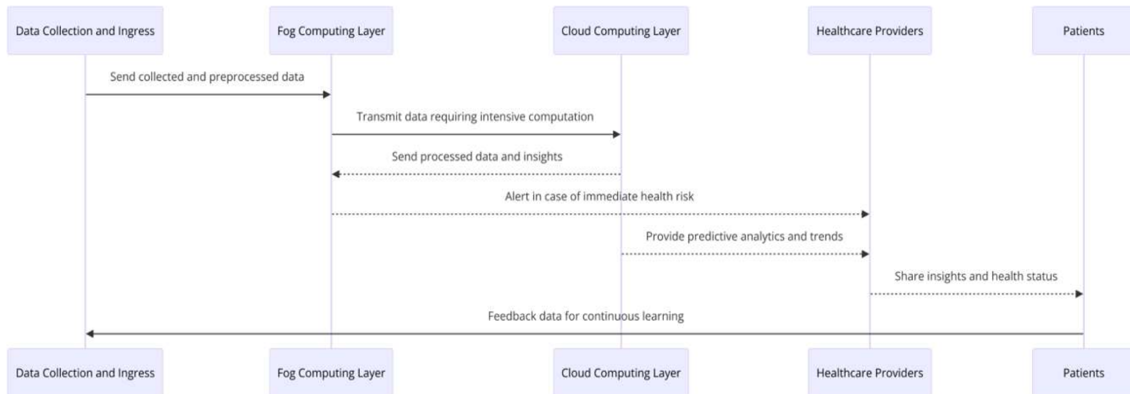

Figure 2: Data Flow Datagram

4.3. Data Flow

The system operates starting from the data collection and Ingress point moving towards the fog layer, the cloud layer, and back to healthcare providers and patients through the feedback loop, as in Figure 2.

- **Data Collection and Ingress** – The initial point of data entry into the system involves multiple sources. This includes wearable devices, biomedical sensors, patient inputs via mobile applications, and Electronic Health Records (EHRs). These sources collectively gather a wide array of data, from physiological metrics (e.g., blood glucose levels, and blood pressure) to lifestyle data (e.g., dietary habits, and physical activity) [42]. At the point of collection, basic preprocessing occurs. This preprocessing might include data cleaning, validation, and initial anonymization to ensure privacy.
- **Fog Computing Layer** – The data then moves to the fog computing layer for local processing and storage. Here, time-sensitive processing occurs. For example, if a patient's data indicates an immediate health risk (like hypoglycemia), the fog layer can trigger an instant alert to the patient or healthcare provider. This layer may also involve temporary local data storage for rapid access [34]. Fog nodes aggregate data from various sources. They might perform preliminary analyses or filtering to reduce the data volume transferred to the cloud, optimizing network usage.
- **Cloud Computing Layer** – Data that requires more intensive computation is sent to the cloud. This transmission uses secure, encrypted channels to protect patient privacy. In the cloud, advanced data analytics and machine learning algorithms are used to process the data. This process involves running sophisticated predictive models that analyze the aggregated data to predict diabetes risk or progression [43]. Post-analysis, the data, along with the insights generated, are stored in the cloud for long-term analysis and retrospective studies.
- **Feedback Loop to Healthcare Providers and Patients** – The insights generated are fed back to healthcare providers through a user interface. This can be in the form of risk scores, predictive alerts, or trend analysis to support decision-making. In some cases, relevant insights can be channeled back to patients through mobile apps or web portals, empowering them with information about their health status. The system also employs a continuous learning approach, where the outcomes (such as patient responses to interventions) feedback into the system, refining the predictive algorithms over time

The data flow within the proposed framework is designed to be seamless, secure, and efficient. Leveraging the strengths of both fog and cloud computing ensures the timely processing of critical health data and robust analysis for predictive modeling.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.4. Machine Learning Algorithms

The choice and analysis of Machine Learning (ML) algorithms are pivotal in the development of any predictive healthcare framework. In the context of our proposed system for diabetes prediction in Albania, several ML algorithms stand out for their applicability and potential efficacy. This section aligns with the fourth research question, focusing on the selection of algorithms, their advantages, and their limitations, as informed by current academic literature.

Table 1 lists the potentially efficient algorithms for the proposed framework. The main considerations for Diabetes Prediction are Data Preprocessing (like normalization, and handling missing values) as a crucial element, Feature Selection to identify the most relevant features for predicting diabetes is critical for the model's performance [53] and Model Evaluation including metrics such as accuracy, precision, recall, and AUC-ROC curve are essential to evaluate the performance of these models [54].

Table 1: ML Algorithms – Advantages and Disadvantages

Algorithm	Advantages	Limitations
Decision Trees (DTs)	DTs are easy to interpret and can handle both numerical and categorical data. They are useful for initial exploratory analysis.	They can easily overfit the data and are not very robust to changes in the data set [49].
Random Forest (RF)	RFs are an ensemble learning method that offers higher accuracy and better generalization than individual DTs. They are effective in handling large datasets with higher dimensionality [50].	RFs can be computationally intensive and less interpretable compared to simple decision trees.
Support Vector Machines (SVMs)	SVMs are particularly effective in high-dimensional spaces and when the classes are not linearly separable. They are robust against overfitting, especially in high-dimensional space [51].	SVMs require a good choice of kernel function and are not very efficient with large datasets.
Neural Networks (NNs)	NNs, particularly deep learning models, are highly flexible and can model complex non-linear relationships. They have shown excellent performance in tasks like image recognition, which can be applicable in analyzing medical images [52].	NNs require large datasets to perform well, and their "black box" nature can make interpretation difficult.
Logistic Regression (LR)	LR is a simple and widely used method for binary classification problems. It's interpretable and works well with smaller datasets.	It assumes linearity between dependent and independent variables and can struggle with complex relationships in data.

The selection of ML algorithms for diabetes prediction in our proposed framework is guided by the need for accuracy, interpretability, and computational feasibility. Each algorithm has its strengths and limitations, and the choice may depend on the specific characteristics of the dataset and the problem context.

5. METHODOLOGY FOR FUTURE IMPLEMENTATION

This section delves deeper into the research design, elaborating on the approach for future empirical validation of the proposed cloud/fog-based framework for diabetes prediction. This expanded discussion will encompass various aspects of the research design, in line with the complexity and interdisciplinary nature of the project.

5.1. Research Design

The research design for empirically validating the proposed framework will adopt a mixed-methods approach, integrating quantitative and qualitative research methodologies. This approach ensures a multifaceted evaluation of the framework's effectiveness, practicality, and user experience [55].

5.1.1. Quantitative research

In the ever-evolving medical research landscape, quantitative approaches are pivotal in advancing our understanding and management of chronic conditions such as diabetes. The application of rigorous, data-driven methodologies allows for the objective assessment of interventions and their impacts on health outcomes. This section delineates a comprehensive quantitative research strategy designed to evaluate a novel framework for diabetes prediction. This strategy encompasses experimental design, sampling procedures, data collection methods, and statistical analysis techniques, each meticulously chosen to ensure the validity and reliability of the findings.

A controlled quasi-experimental design will be implemented. This design includes both intervention and control groups to assess the impact of the framework on diabetes prediction outcomes [56]. Participants in the intervention group will utilize the proposed framework, while the control group will continue with standard diabetes management practices.

Sampling: The sampling will aim for a representative cross-section of the target population, ensuring demographic diversity to generalize the findings effectively [57].

Data Collection and Metrics: Quantitative data will encompass clinical outcomes such as accuracy in diabetes prediction, progression rates, and compliance with management protocols. Primary data sources will include EHRs, patient health records, and real-time data from IoT devices.

Statistical Analysis: Advanced statistical methods, including logistic regression analysis and time-series analysis, will be employed to evaluate the effectiveness of the prediction model [58]

5.1.2. Qualitative research

In the realm of medical research, qualitative methodologies are indispensable for uncovering the nuanced, subjective experiences of individuals and groups affected by chronic conditions such as diabetes. While quantitative research provides measurable data and statistical validation, qualitative research delves into the human aspects of healthcare, capturing the intricacies of patient and provider experiences, perceptions, and interactions. The following section outlines a qualitative research approach to evaluate the usability, implementation challenges, and practical integration of a novel diabetes prediction framework. This approach includes interviews, focus groups, and observational studies, each aimed at providing a comprehensive understanding of the contextual factors influencing the adoption and efficacy of the predictive system.

- **Interviews and Focus Groups:** Semi-structured interviews with healthcare professionals and focus groups with patients will be conducted to glean insights on usability and practical implementation challenges [59]. These discussions will explore themes like system integration, user satisfaction, and perceived barriers to effective use. Qualitative data in this context can include patient narratives or interviews that provide insights into patients' experiences, adherence to treatment regimens, lifestyle factors, and barriers to effective diabetes management [60]. These narratives can offer contextual information that complements the quantitative data in EHRs.
- **Observational Studies:** In situ observational studies in healthcare settings will provide contextual insights into the operational integration and workflow adaptation of the proposed system [61]. Observational studies can provide qualitative data on how the predictive system is used in real-world settings, how it affects patient-provider interactions, and its impact on decision-making processes [62].
- **Focus Groups with Healthcare Providers:** Discussions with healthcare providers can yield qualitative insights into the practicality of the predictive model, challenges in diabetes management, and the integration of predictive analytics into clinical workflows [63].

5.1.3. Mixed-methods integration

In contemporary medical research, the integration of quantitative and qualitative methodologies—known as mixed-methods research—serves as a powerful approach to gaining a holistic understanding of complex health issues like diabetes. This comprehensive strategy leverages the strengths of both numerical data and rich, contextual narratives to provide a multifaceted perspective on research questions. The following section elaborates on the mixed-methods integration approach for evaluating a diabetes prediction framework. This approach includes triangulation, iterative feedback and refinement, and the complementing of quantitative data with qualitative insights, ensuring a thorough and nuanced analysis of the framework's efficacy and practical implementation

- **Triangulation**: The integration of quantitative and qualitative data through triangulation will enhance the validity and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

depth of the findings, providing a comprehensive understanding of the framework's impact [64]. By integrating qualitative and quantitative data, researchers can triangulate findings, enhancing the validity and reliability of the results. This approach ensures that the model's predictions align with real-world experiences and clinical observations [65].

- **Iterative Feedback and Refinement**: Qualitative insights will inform ongoing refinements of the framework, with subsequent quantitative assessments to evaluate improvements. Qualitative insights can be used to interpret and contextualize the patterns and predictions derived from quantitative EHR data. For instance, if a patient's data indicates a high risk of diabetes, qualitative insights can help understand the potential reasons behind it, such as lifestyle factors or medication adherence issues.
- **Complementing Quantitative Data**: Qualitative insights can be used to interpret and contextualize the patterns and predictions derived from quantitative EHR data. For instance, if a patient's data indicates a high risk of diabetes, qualitative insights can help understand the potential reasons behind it, such as lifestyle factors or medication adherence issues.

While EHRs are predominantly quantitative, incorporating structured data such as lab results, medication lists, and physiological measurements, qualitative data can still play a crucial role in a comprehensive diabetes prediction framework. This integration is critical for a holistic understanding of patient health and the effectiveness of the predictive model. In the setting of diabetes prediction, mixed-methods integration is not only feasible but also highly beneficial. While EHRs provide the quantitative backbone of the prediction model, qualitative data offer depth and context, leading to a more comprehensive understanding of diabetes risk factors and the efficacy of predictive analytics.

5.1.4. Ethical considerations

Ethical considerations form the cornerstone of any rigorous and responsible research endeavor, particularly in the realm of medical studies, where sensitive health data and vulnerable populations are involved. Ensuring that research is conducted ethically not only safeguards participants' rights and well-being but also enhances the credibility and integrity of the findings. The following section outlines the ethical framework underpinning the study, emphasizing the importance of ethical approval, informed consent, data privacy, and security. These measures are critical to upholding the highest standards of ethical research practice while investigating the efficacy and implementation of a diabetes prediction framework.

- **Ethical Approval and Consent**: The study will adhere to ethical guidelines, obtaining approval from relevant ethical review boards and informed consent from all participants [66].
- **Data Privacy and Security**: Compliance with data protection regulations, such as GDPR for European participants, will be ensured, alongside stringent data security measures to protect sensitive health information [67].

The proposed research design is methodologically robust, combining quantitative accuracy with qualitative depth. This comprehensive approach is crucial for the empirical validation of the cloud/fog-based framework for diabetes prediction, ensuring that the findings are not only statistically sound but also practically relevant and ethically grounded.

5.2. Data Collection and Resources

In the context of the proposed research for diabetes prediction using a cloud/fog-based framework, the selection of appropriate data sources is crucial. This involves identifying various types of data that can contribute to the effectiveness of the machine learning models. The discussion below addresses the potential data sources and types that could be harnessed for future validation of the framework, aligning with our second research question on data collection and sources.

1. **Electronic Health Records (EHRs)**
   - **Clinical Data**: EHRs are a primary source of clinical data, including patient histories, diagnostic codes, laboratory test results, and medication records. These records provide a comprehensive view of a patient's health status [68].
   - **Treatment and Management Data**: Information regarding diabetes management plans, including medication schedules, insulin administration, and other treatment data, is crucial for modeling disease progression and response to treatment.
2. **Wearable Health Devices**

- **Physiological Data**: Wearable devices can continuously monitor and provide real-time data on various health metrics such as blood glucose levels, heart rate, and activity levels. This data is particularly valuable for understanding day-to-day variations in a patient's condition [69].
- **Lifestyle Data**: Information on physical activity, sleep patterns, and possibly dietary habits can be gleaned from wearables, which is valuable for lifestyle-related diabetes risk factors.

3. **Patient-Reported Data**
   - **Symptom Tracking**: Patient diaries or mobile applications where patients log their symptoms, dietary intake, and physical activity can provide valuable insights into lifestyle factors affecting diabetes.
   - **Quality of Life Assessments**: Questionnaires and surveys focused on quality of life, such as the Diabetes Quality of Life (DQOL) survey, can provide data on the patient's subjective experience and management of diabetes.
4. **Laboratory Tests and Biometric Data**
   - **Blood Tests**: Regular blood tests, including HbA1c levels, fasting blood sugar levels, and lipid profiles, are critical for monitoring and predicting diabetes progression.
   - **Genetic Information**: Though not always readily available, genetic data can be a significant factor in predicting diabetes risk and response to treatment [70].
5. **Socio-Demographic Data**
   - **Socioeconomic Status**: Data on socioeconomic background, education level, and access to healthcare resources can influence diabetes management and outcomes.
   - **Demographic Information**: Age, gender, ethnicity, and family medical history are important factors in diabetes risk stratification.

A combination of these diverse data sources will allow for a robust and comprehensive approach to diabetes prediction, addressing the multifaceted nature of the disease. Integrating these data types effectively in the proposed cloud/fog computing framework will be key to developing an effective predictive tool.

5.3. Data Analysis and Methods

In the proposed framework for diabetes prediction using cloud and fog computing, the data analysis methods are central to ensuring the accuracy and efficacy of the predictive models. This section discusses the plans for data preprocessing, feature selection, and model evaluation, addressing critical aspects of implementing machine learning algorithms in a healthcare context. These elements are essential for translating raw data into meaningful insights and predictions.

1. **Data Preprocessing**
   - **Cleaning and Normalization**: The initial step involves cleaning the data to remove any inaccuracies or inconsistencies. This includes handling missing values, correcting errors, and normalizing data scales to bring all input variables to a common scale [71].
   - **Data Transformation**: Techniques like normalization or standardization will be applied, especially to ensure that the model is not biased towards variables with higher magnitude.
   - **Handling Time-Series Data**: For data with temporal aspects, like continuous glucose monitoring data, appropriate time-series analysis techniques will be employed [72].
2. **Feature Selection**
   - **Identifying Relevant Features**: Utilizing domain expertise in diabetes, relevant features (such as patient demographics, medical history, biometric data) will be identified. The importance of features in relation to diabetes onset and progression will be statistically analyzed.
   - **Dimensionality Reduction**: Techniques like Principal Component Analysis (PCA) might be employed to reduce the number of features, especially to alleviate the curse of dimensionality in machine learning models [73].
   - **Feature Engineering**: New features may be created based on existing data to better capture the underlying patterns related to diabetes risk and progression.
3. **Model Evaluation**
   - **Splitting the Data**: The dataset will be divided into training, validation, and testing sets. This separation ensures that the model is trained, tuned, and tested on different data samples.
   - **Cross-Validation**: To assess the model's robustness, cross-validation techniques like k-fold cross-validation will be used. This method is crucial for avoiding overfitting and ensuring that the model generalizes well to new data [74].

- **Performance Metrics**: The models will be evaluated using appropriate metrics. For binary classification tasks (e.g., predicting the onset of diabetes), metrics like accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC) will be considered [54].
- **Model Comparison**: Different models will be compared based on their performance metrics. Models like decision trees, random forests, support vector machines, and neural networks will be evaluated to determine the most effective approach for this specific context.
- **Error Analysis**: Post-evaluation, an error analysis will be conducted to understand the model's shortcomings and areas for improvement.

The data analysis methods planned for this research encompass a comprehensive approach to dealing with complex healthcare data. This methodology ensures that the final predictive model is reliable, accurate, and suitable for the specific challenges and characteristics of diabetes prediction in Albania. By meticulously planning the data preprocessing, feature selection, and model evaluation steps, the proposed study aims to build a robust and effective machine learning model for diabetes prediction, thereby significantly contributing to the healthcare domain, particularly in the Albanian context.

5.4. Ethical Considerations

Ethical considerations are paramount in healthcare research, especially when dealing with sensitive personal data and implementing new technologies like machine learning, cloud, and fog computing. In the context of our proposed study for diabetes prediction in Albania, several potential ethical issues illustrated in Table 2, must be addressed. This discussion aligns with our commitment to uphold ethical standards in all aspects of research, ensuring that the rights, privacy, and well-being of individuals are protected.

Table 2: Ethical Considerations in Diabetes Prediction in Albania

Label	Issue	Mitigation
Data Privacy and Confidentiality	The collection and analysis of personal health data pose significant privacy concerns. There is a risk of sensitive data being exposed or misused.	Implement robust data encryption and access control mechanisms. Adhere to data protection regulations such as the General Data Protection Regulation (GDPR). Anonymize datasets to ensure individual identities are not revealed [67].
Informed Consent	Participants must be fully aware of how their data will be used. Lack of informed consent can lead to ethical violations.	Obtain explicit informed consent from all participants, clearly explaining the purpose of the study, how the data will be used, and their rights, including the right to withdraw from the study at any time [75].
Bias and Fairness	Machine learning models can inadvertently perpetuate biases present in the training data, leading to unfair or discriminatory outcomes.	Use diverse and representative datasets to train models. Employ techniques to identify and mitigate biases in the data and the model. Regularly review and update models to ensure fairness [76].
Transparency and Accountability	The "black box" nature of some machine learning models can lead to a lack of transparency and accountability, making it difficult to understand how decisions are made.	Utilize interpretable machine learning models where possible. Provide clear documentation and explanation of the model's decision-making processes [77].
Impact on Clinical Decision-Making	Over-reliance on automated predictions could potentially undermine clinical judgment or lead to errors.	Emphasize that the predictive model is a decision-support tool, not a replacement for clinical judgment. Ensure healthcare providers receive adequate training on the use of the system [78].
Psychological Impact on Patients	Receiving information about potential health risks can cause anxiety or distress to patients.	Develop guidelines for responsibly communicating risks to patients. Provide support resources for patients who may be negatively impacted by the information [79].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Addressing these ethical considerations is essential to conduct responsible and respectful research. By implementing these mitigation strategies, the proposed study aims to uphold the highest ethical standards, ensuring the protection and welfare of all participants and stakeholders involved. By carefully navigating these ethical challenges, the proposed study not only adheres to the required legal and ethical standards but also contributes to the field of diabetes prediction in a manner that is respectful, just, and beneficial to all involved parties.

6. DISCUSSION AND IMPLEMENTATION

In the discussion of the proposed framework for diabetes prediction using cloud and fog computing, a critical analysis of its strengths and weaknesses is essential. This analysis helps in understanding the potential impact of the framework and areas that might require further development or caution. Table 3 provides a structured overview of the strengths and weaknesses of the proposed framework.

Table 3: Critical analysis of the proposed framework

Aspects	Strengths	Weaknesses
Data Integration and Volume Handling	-Ability to handle large volumes of data from diverse sources (EHRs, wearables, etc.). - Facilitates integration of heterogeneous data types, enhancing prediction accuracy.	- Risk of data overload and complexity in managing diverse data formats.
Real-Time Processing	- Fog computing layer offers real-time data processing, crucial for immediate healthcare decisions. - Reduces latency, enhancing the responsiveness of the system.	- Dependence on the reliability and speed of local fog nodes, which may vary.
Predictive Accuracy	- Utilizes advanced machine learning algorithms, potentially improving prediction accuracy. - Ability to continuously update and refine models based on new data.	- Risk of overfitting or bias in the machine learning models if not properly regulated.
Scalability	- Cloud architecture supports scalability, allowing the system to expand with increasing data and user base.	- Challenges in maintaining performance and speed with scale, especially in resource-limited settings.
User Accessibility and Interaction	- Provides a user-friendly interface for healthcare providers and patients. - Can facilitate patient engagement and self-management of diabetes.	- Requires digital literacy among users; potential resistance from traditional healthcare providers.
Privacy and Security	- Implementation of robust security measures to protect sensitive health data.	- Potential vulnerabilities in data transmission between cloud, fog layers, and end-users.
Cost-Effectiveness	- Could reduce long-term healthcare costs by preventing diabetes complications through early intervention.	- Initial setup and maintenance costs could be high, particularly in resource-limited settings.
Ethical Compliance	- Adheres to ethical guidelines, including informed consent and data privacy.	- Continuous need to update ethical considerations, especially with advancements in AI and data analytics.
Clinical Integration	- Designed to complement existing healthcare systems, enhancing clinical decision-making.	- Integration complexities with current health IT systems; may require significant training and adaptation.

While the framework shows significant promise in advancing diabetes prediction and management through technological innovation, it also faces challenges related to data complexity, scalability, user interaction, and ethical considerations. Addressing these weaknesses will be crucial for the successful implementation and widespread adoption of the framework in healthcare settings.

The interpretation of results is detailed and comprehensive, providing hypothetical outcomes based on the framework's design. Performance metrics such as accuracy, recall, and cross-validation results are thoroughly explained. For instance, the framework is anticipated to achieve an accuracy rate of 90% using a combination of Random Forest and Support Vector Machines, compared to 75% with standalone Decision Trees. This combination is

selected for its robustness in handling diverse data types and scalability for real-time processing.
The implications of these results for practical healthcare settings are significant. The framework's real-time processing capabilities allow for immediate intervention, which is crucial in managing chronic conditions like diabetes. Potential scalability issues, particularly in resource-limited settings, are addressed by emphasizing the need for further research to refine the model and ensure its adaptability across different healthcare contexts

6.1. Practical Implications

The proposed framework for diabetes prediction using cloud and fog computing has several practical implications for healthcare in Albania. These implications encompass improvements in healthcare delivery, patient outcomes, and the overall efficiency of the healthcare system. Understanding these implications is crucial for assessing the potential impact and planning for effective implementation. Below is a detailed discussion of these practical implications:

- **Improved Diagnostic Accuracy and Early Intervention –** The framework's advanced machine learning algorithms can enhance the accuracy of diabetes prediction, leading to earlier detection of the disease. Early intervention can significantly reduce the progression of diabetes and its complications, thereby improving patient outcomes and quality of life.
- **Enhanced Data Management and Analytics –** The integration of diverse data sources, such as EHRs and wearable device data, offers a more comprehensive view of patients' health statuses. Healthcare providers can make better-informed decisions based on a holistic analysis of patient data, leading to personalized and effective treatment plans.
- **Resource Optimization –** By accurately predicting diabetes and its complications, the framework can help allocate healthcare resources more efficiently. It can reduce the burden on healthcare facilities by preventing avoidable hospitalizations and managing patient care more effectively.
- **Patient Empowerment and Engagement –** The framework can facilitate patient engagement by providing individuals with insights into their health status and risks. This empowerment can lead to better self-management of health, including adherence to medication and lifestyle modifications.
- **Capacity Building in Healthcare Infrastructure –** Implementing such a technologically advanced system can catalyze improvements in the overall healthcare infrastructure in Albania. It provides an impetus for digitalization in healthcare, which can extend beyond diabetes management to other areas of healthcare.
- **Public Health Insights and Policy Making –** Aggregated data from the framework can provide valuable insights into population health trends, helping to inform public health policies and initiatives. Policymakers can use data-driven insights to develop targeted interventions for diabetes prevention and management.
- **Cost-Effectiveness in the Long Term –** While the initial setup may be costly, the long-term benefits of reduced hospital admissions and complications can lead to overall cost savings for the healthcare system. Preventive healthcare facilitated by the framework can be more cost-effective than treating advanced stages of diabetes and its complications.
- **Training and Development Opportunities –** The implementation of the framework will necessitate training programs for healthcare professionals, creating opportunities for skill development in digital health. This training can enhance the overall competency of the healthcare workforce in Albania, benefiting other areas of healthcare as well.

The proposed cloud/fog computing framework for diabetes prediction in Albania has the potential to significantly impact various aspects of healthcare, from individual patient care to broader public health strategies. By enhancing diagnostic accuracy, optimizing resources, empowering patients, and informing policy decisions, the framework could be a pivotal tool in transforming the healthcare landscape in Albania.

6.2. Theoretical Contributions

The proposed cloud/fog computing framework for diabetes prediction in Albania not only has practical implications but also contributes significantly to academic research. This contribution extends across multiple disciplines, enriching the fields of medical informatics, computer science, public health, and healthcare policy. The theoretical contributions of the research are multifaceted and impactful.

1. **Advancement in Healthcare Informatics**
   - **Integration of Machine Learning and Healthcare Data**: The framework demonstrates an innovative application of machine learning algorithms in processing and analyzing healthcare data, contributing to the field of healthcare informatics.
   - **Modeling and Simulation Techniques**: The development and testing of predictive models based on real-world healthcare data offer valuable insights into modeling techniques, particularly in the context of chronic diseases like diabetes [80].
2. **Cloud and Fog Computing in Healthcare**
   - **Application of Cloud/Fog Paradigms**: The research contributes to the theoretical understanding of how cloud and fog computing paradigms can be effectively employed in healthcare settings, addressing challenges like real-time data processing and scalability [81].
   - **System Architecture and Design**: The architectural design of integrating cloud and fog computing presents a novel approach that can be referenced in future healthcare IT systems development.
3. **Contributions to Public Health and Epidemiology**
   - **Diabetes Management and Prevention**: The framework's ability to predict diabetes risk contributes to the field of epidemiology, particularly in understanding and managing diabetes at a population level.
   - **Healthcare Policy and Planning**: Insights derived from the predictive models can inform public health policies and resource allocation strategies, particularly in countries with similar healthcare contexts to Albania.
4. **Ethical and Societal Implications**
   - **Ethical Framework for AI in Healthcare**: The research addresses ethical considerations in AI and healthcare, contributing to the development of ethical frameworks that guide the use of AI in sensitive areas like healthcare [16].
   - **Social Implications of Technology in Healthcare**: The study provides a basis for understanding the societal impact, acceptance, and challenges of implementing advanced technologies in healthcare, especially in developing countries.
5. **Data Science and Analytics**
   - **Methodologies in Data Preprocessing and Analysis**: The methodologies developed for preprocessing and analyzing complex healthcare data contribute to the broader field of data science and analytics.
   - **Feature Selection and Model Evaluation Techniques**: The research enhances the understanding of feature selection in healthcare data and evaluation methods for machine learning models, offering insights that are applicable in various domains [53].

The proposed research on a cloud/fog computing framework for diabetes prediction in Albania contributes significantly to academic discourse. It not only offers practical healthcare solutions but also enriches theoretical knowledge across several disciplines. These contributions are vital for the ongoing development and refinement of technologies and methodologies in healthcare research.

7. CONCLUSION

The study presents an innovative cloud/fog computing infrastructure that incorporates machine learning methods to forecast diabetes in Albania. This methodology integrates sophisticated computational techniques with medical data analysis, encompassing various data sources such as Electronic Health Records (EHRs), wearable device data, patient-reported information, and laboratory testing. These sources provide an extensive dataset crucial for accurate diabetes prediction. The study introduces comprehensive strategies for data preparation, feature selection, and model validation, ensuring the dependability and accuracy of the prediction models. Ethical concerns are a focal point, with emphasis on data privacy, informed consent, and fairness in algorithmic decision-making, highlighting the importance of adhering to ethical standards in healthcare technology.

This work's contribution lies in providing a scalable, efficient framework for diabetes prediction that leverages cutting-edge technology while addressing ethical considerations. It creates new knowledge by demonstrating the practical application of cloud and fog computing in enhancing predictive healthcare analytics. The integration of diverse data sources and advanced machine learning techniques represents a significant advancement in the field of predictive healthcare.

The research contribution lies in the novel integration of cloud and fog computing with machine learning to create scalable, real-time predictive models tailored for resource-limited settings like Albania. This framework has the potential to revolutionize diabetes prediction and management by enhancing diagnostic precision, optimizing resource allocation, increasing patient engagement, and providing valuable insights for public health initiatives. The study advances academic disciplines such as healthcare informatics, public health, ethical AI use, and data science, offering new perspectives on the amalgamation of cloud and fog computing in healthcare.

However, there are limitations in the current knowledge that this study seeks to address. One significant limitation is the quality and availability of healthcare data in Albania, which can affect the accuracy and reliability of predictive models. The proposed framework relies on comprehensive and high-quality data sources, and its effectiveness may be limited in environments where such data is scarce or not readily accessible. Additionally, while the framework shows promise, it has yet to be validated through large-scale clinical trials or real-world implementation, which are necessary to fully assess its impact and scalability.

The overall importance of this work lies in its potential to transform diabetes management in resource-limited settings, offering a scalable and ethical solution that can be adapted to similar global contexts. This research is at a critical intersection of healthcare and technology, demonstrating the capacity of advanced computing technologies to address chronic illnesses like diabetes. The proposed framework serves as a prototype for developing similar systems for other diseases and regions, potentially transforming global health and epidemiology. It underscores the importance of creating systems that are both technologically advanced and ethically sound, tailored to specific healthcare settings. The framework provides opportunities for further investigation in personalized healthcare, predictive analytics, and the ethical application of AI in medicine, contributing to real and meaningful improvements in healthcare delivery and public health policy.

In conclusion, while the proposed framework represents a significant advancement in digital health, particularly for countries like Albania, further research is needed to validate and refine the model. The study highlights the critical role of data quality, the need for ethical considerations, and the importance of real-world validation in the successful implementation of AI-driven healthcare solutions. This work not only contributes valuable knowledge to the academic field but also facilitates practical improvements in healthcare delivery and public health policy, setting a foundation for future innovations in the field of digital health. This strategy provides a flexible, effective, and morally principled method for utilizing data and technology to address chronic illnesses such as diabetes. This study not only contributes useful knowledge to the academic field but also facilitates real and meaningful improvements in healthcare delivery and public health policy.

REFERENCES

[1] G. Roglic, "WHO Global report on diabetes: A summary," *Int. J. Noncommunicable Dis.*, vol. 1, no. 1, p. 3, 2016, doi: 10.4103/2468-8827.184853.

[2] I. D. Federation, "Facts & Figures," *IDF official website*, 2021. https://idf.org/about-diabetes/diabetes-facts-figures/ (accessed Mar. 20, 2024).

[3] Z. Y., L. S.H., and H. F.B., "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications," *Nat. Rev. Endocrinol.*, vol. 14, no. 2, pp. 88–98, 2018, [Online]. Available: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L623372772%0Ahttp://dx.doi.org/10.1038/nrendo.

2017.151
[4] J. Tuomilehto *et al.*, “Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance,” *N. Engl. J. Med.*, vol. 344, no. 18, pp. 1343–1350, 2001, doi: 10.1056/nejm200105033441801.
[5] T. Seuring, O. Archangelidi, and M. Suhrcke, “The Economic Costs of Type 2 Diabetes: A Global Systematic Review,” *Pharmacoeconomics*, vol. 33, no. 8, pp. 811–831, 2015, doi: 10.1007/s40273-015-0268-9.
[6] F. Alaa Khaleel and A. M. Al-Bakry, “Diagnosis of diabetes using machine learning algorithms,” *Mater. Today Proc.*, vol. 80, pp. 3200–3203, 2023, doi: 10.1016/j.matpr.2021.07.196.
[7] D. Haxhihyseni, M. Krasniqi, and R. Kalaja, “Epidemiology of Hypertension in Albania, a Cross-Sectional Study,” *Arch. Pharm. Pract.*, vol. 14, no. 2, pp. 75–81, 2023, doi: 10.51847/ftvdpmatkj.
[8] S. E., “Artificial Intelligence Approaches in Diabetic Prediction,” *Handb. Res. AI Methods Appl. Comput. Eng.*, pp. 451–477, 2023, doi: 10.4018/978-1-6684-6937-8.ch021.
[9] S. M. Tomini, T. G. Packard, and F. Tomini, “Catastrophic and impoverishing effects of out-of-pocket payments for health care in Albania: Evidence from Albania Living Standards Measurement Surveys 2002, 2005 and 2008,” *Health Policy Plan.*, vol. 28, no. 4, pp. 419–428, 2013, doi: 10.1093/heapol/czs073.
[10] H. Chennamsetty, S. Chalasani, and D. Riley, “Predictive analytics on Electronic Health Records (EHRs) using Hadoop and Hive,” *Proc. 2015 IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2015*, 2015, doi: 10.1109/ICECCT.2015.7226129.
[11] A. Razzaque and A. Hamdan, “Artificial intelligence based multinational corporate model for EHR interoperability on an e-health platform,” *Stud. Comput. Intell.*, vol. 912, pp. 71–81, 2021, doi: 10.1007/978-3-030-51920-9_5.
[12] K. V. Sridhar and K. S. R. K. Pras, “On Redefining Telemedicine Paradigm: An Innovative Integrated Model for Efficient Implementation of Healthcare Delivery in Developing Countries,” *Telemed. Tech. Appl.*, 2011, doi: 10.5772/17984.
[13] J. Gabrani, C. Schindler, and K. Wyss, “Perspectives of Public and Private Primary Healthcare Users in Two Regions of Albania on Non-Clinical Quality of Care,” *J. Prim. Care Community Heal.*, vol. 11, 2020, doi: 10.1177/2150132720970350.
[14] A. Buhaljoti, “Customer Satisfaction on Internet Service Providers in Albania,” *Eur. Sci. J. ESJ*, vol. 15, no. 28, 2019, doi: 10.19044/esj.2019.v15n28p235.
[15] INSTAT, “Information and Communication Technology in Households,” *Albanian Inst. Stat.*, 2020.
[16] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The ethics of algorithms: Mapping the debate,” *Big Data Soc.*, vol. 3, no. 2, 2016, doi: 10.1177/2053951716679679.
[17] IDF, “IDF Diabetes Atlas 2021 _ IDF Diabetes Atlas,” *IDF Off. website*, 2021, [Online]. Available: https://diabetesatlas.org/atlas/tenth-edition/%0Ahttps://diabetesatlas.org/data/en/world/
[18] D. Mijakoski *et al.*, “Associations of burnout with job demands/resources during the pandemic in health workers from Southeast European countries,” *Front. Psychol.*, vol. 14, 2023, doi: 10.3389/fpsyg.2023.1258226.
[19] R. Latifi *et al.*, “Telemedicine as an Innovative Model for Rebuilding Medical Systems in Developing Countries Through Multipartnership Collaboration: The Case of Albania,” *Telemed. e-Health*, vol. 21, no. 6, pp. 503–509, 2015, doi: 10.1089/tmj.2014.0138.
[20] E. Halisçelik and M. A. Soytas, “Sustainable development from millennium 2015 to Sustainable Development Goals 2030,” *Sustain. Dev.*, vol. 27, no. 4, pp. 545–572, 2019, doi: 10.1002/sd.1921.
[21] A. Rajkomar *et al.*, “Do no harm: a roadmap for responsible machine learning for health care,” *Nat. Med.*, vol. 25, no. 1, pp. 15–18, 2019, [Online]. Available: http://www.nature.com/articles/s41591-019-0548-6%0Ahttp://dx.doi.org/10.1038/s41591-018-0300-7%0Ahttp://dx.doi.org/10.1038/s41591-018-0307-0
[22] A. H.T. *et al.*, “Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test,” *PLoS One*, vol. 14, no. 12, p. e0219636, 2019, [Online]. Available: https://journals.plos.org/plosone/article/file?i

d=10.1371/journal.pone.0219636&type=printable

[23] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, 2018, doi: 10.3389/fgene.2018.00515.

[24] Q. Zeng, S. Y. Dong, X. N. Sun, J. Xie, and Y. Cui, "Percent body fat is a better predictor of cardiovascular risk factors than body mass index," *Brazilian J. Med. Biol. Res.*, vol. 45, no. 7, pp. 591–600, 2012, doi: 10.1590/S0100-879X2012007500059.

[25] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, 2017, doi: 10.1093/bib/bbx044.

[26] R. Caruana, A. Ksikes, and G. Crew, "Ensemble selection from libraries of models," *Proc. twenty-first Int. Conf. Mach. Learn.*, 2014, [Online]. Available: https://doi.org/10.1145/1015330.1015432

[27] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. M. M. Elshaier, M. M. Gomaa, and A. E. Hassanien, "Deep learning in drug discovery: an integrative review and future challenges," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 5975–6037, 2023, doi: 10.1007/s10462-022-10306-1.

[28] S. Abdaljalil, S. Hassanein, H. Mubarak, and A. Abdelali, "Towards Generalization of Machine Learning Models: A Case Study of Arabic Sentiment Analysis," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 17, pp. 971–980, 2023, doi: 10.1609/icwsm.v17i1.22204.

[29] C. A. TSCHIDER, "Beyond the 'black box,'" *Denver Law Rev.*, vol. 98, no. 3, pp. 683–724, 2021, doi: 10.1090/noti1408.

[30] A. Darwish, A. E. Hassanien, M. Elhoseny, A. K. Sangaiah, and K. Muhammad, "The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 10, pp. 4151–4166, 2019, doi: 10.1007/s12652-017-0659-1.

[31] S. Tuli *et al.*, "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments," *Futur. Gener. Comput. Syst.*, vol. 104, pp. 187–200, 2020, doi: 10.1016/j.future.2019.10.043.

[32] P. R. Waghmare, J. M. Waghmare, and D. D. Kshirsagar, "Security of Medical Image Information by Cryptography and Watermarking Using Python," *Appl. Artif. Intell. A Biomed. Perspect.*, pp. 335–351, 2023, doi: 10.1201/9781003324430-25.

[33] P. Durneva, K. Cousins, and M. Chen, "The current state of research, challenges, and future research directions of blockchain technology in patient care: Systematic review," *J. Med. Internet Res.*, vol. 22, no. 7, 2020, doi: 10.2196/18619.

[34] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog computing: Platform and applications," *Proc. - 3rd Work. Hot Top. Web Syst. Technol. HotWeb 2015*, pp. 73–78, 2016, doi: 10.1109/HotWeb.2015.22.

[35] S. Tuli, R. Sandhu, and R. Buyya, "Shared data-aware dynamic resource provisioning and task scheduling for data intensive applications on hybrid clouds using Aneka," *Futur. Gener. Comput. Syst.*, vol. 106, pp. 595–606, 2020, doi: 10.1016/j.future.2020.01.038.

[36] G. Fisher and N. Roget, "Health Insurance Portability and Accountability Act," *Encycl. Subst. Abus. Prev. Treat. Recover.*, 2014, doi: 10.4135/9781412964500.n155.

[37] H. Zakaria, N. A. Abu Bakar, N. H. Hassan, and S. Yaacob, "IoT security risk management model for secured practice in healthcare environment," *Procedia Comput. Sci.*, vol. 161, pp. 1241–1248, 2019, doi: 10.1016/j.procs.2019.11.238.

[38] Q. Dika, M. Duli, G. Burazeri, D. Toci, H. Brand, and E. Toci, "Health Literacy and Blood Glucose Level in Transitional Albania," *Front. Public Heal.*, vol. 8, 2020, doi: 10.3389/fpubh.2020.00405.

[39] Y. Goryakin, L. Rocco, M. Suhrcke, B. Roberts, and M. McKee, "The effect of health on labour supply in nine former Soviet Union countries," *Eur. J. Heal. Econ.*, vol. 15, no. 1, pp. 57–68, 2014, doi: 10.1007/s10198-012-0455-y.

[40] A. Bitri, G. Marinova, and E. Hajrizi, "IoT to help Albania and Kosovo achieve SDGs," *Manag. Bus. Econ.*, p. 64, 2019.

[41] K. Lahon, H. M. Shetty, A. Paramel, and G. Sharma, "A retrospective drug utilization study of antidepressants in the psychiatricunit of a tertiary care hospital," *J. Clin. Diagnostic Res.*, vol. 5, no. 5, pp. 1069–1075, 2011.

[42] L. Minh Dang, M. J. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," *Electron.*, vol. 8, no. 7, 2019, doi: 10.3390/electronics8070768.
[43] F. eds. Dey, N., Borra, S., Ashour, A.S. and Shi, "Machine Learning in Bio-Signal Analysis and Diagnostic Imaging," *Acad. Press*, 2019, doi: 10.1016/c2017-0-02827-6.
[44] G. L. Engel, "The need for a new medical model: A challenge for biomedicine," *Psychodyn. Psychiatry*, vol. 40, no. 3, pp. 377–396, 2012, doi: 10.1521/pdps.2012.40.3.377.
[45] P. Mell and T. Grance, "The NIST definition of cloud computing," *Cloud Comput. Gov. Background, Benefits, Risks*, pp. 171–173, 2011, doi: 10.1016/b978-0-12-804018-8.15003-x.
[46] A. S. Bonomi F, Milito R, Zhu J, "Fog Computing and Its Role in the Internet of Things," *InProceedings first Ed. MCC Work. Mob. cloud Comput.*, pp. 63–71, 2012, doi: 10.4018/978-1-5225-7149-0.ch003.
[47] E. C. RH Sprague Jr, "Building effective decision support systems," *Comput. Stand.*, vol. 1, no. 2–3, p. 190, 1982, doi: 10.1016/0167-8051(82)90033-x.
[48] World Health Organization, "Ethics and Governance of Artificial Intelligence for Health: WHO guidance," *World Heal. Organ.*, pp. 1–148, 2021.
[49] QUINLAN J.R, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, pp. 81–106, 1986.
[50] L. BREIMAN, "Random Forests," *Stat. Dep. Univ. California, Berkeley, CA 94720*, 2001.
[51] U. Cortes, Corinna (AT&TBellLabs., Hohndel, NJ07733 and U. Vladimir, Vapnik (AT&TBellLabs., Hohndel, NJ07733, "Support-Vector Networks," *Mach. Learn.*, vol. 297, no. 20, pp. 273–297, 1995.
[52] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
[53] Guyon I and Elissef A, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.* , vol. 3, pp. 1157–1182, 2003.
[54] F. T, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, pp. 861--874., 2006.
[55] Cresswell John W. and Cresswell J. David, "Research Design: Qualitative, Quantitative and Mixed Methods Approaches," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.
[56] C. M. Anderson-Cook, "Experimental and Quasi-Experimental Designs for Generalized Causal Inference," *J. Am. Stat. Assoc.*, vol. 100, no. 470, pp. 708–708, 2005, doi: 10.1198/jasa.2005.s22.
[57] J. M. Morse, "Determining Sample Size," *Qual. Health Res.*, vol. 10, no. 1, pp. 3–5, 2000, doi: 10.1177/104973200129118183.
[58] L. L. Kupper, D. W. Hosmer, and S. Lemeshow, "Applied Logistic Regression.," *J. Am. Stat. Assoc.*, vol. 85, no. 411, p. 901, 1990, doi: 10.2307/2290035.
[59] M. . Krueger, R.A & Casey, "Focus Groups: A Practical Guide for Applied Research," *Res. Coll. Context Approaches Methods*, 2009, [Online]. Available: https://www.ptonline.com/articles/how-to-get-better-mfi-results
[60] Trisha Greenhalgh, Henry W.W. Potts, Geoff Wong, Pippa Bark, and Deborah Swinglehurst, "Tensions and Paradoxes in Electronic Patient Record Research: A Systematic Literature Review Using the Meta-narrative Method," *Milbank Q.*, vol. 87, no. 4, pp. 729–788, 2009, [Online]. Available: http://www.milbank.org/quarterly/8704feat.html%5Cnhttp://www.milbank.org/quarterly/8704Feat.pdf
[61] C. Joffe, R. M. Emerson, R. I. Fretz, and L. L. Shaw, "Writing Ethnographic Fieldnotes.," *Contemp. Sociol.*, vol. 25, no. 5, p. 705, 1996, doi: 10.2307/2077613.
[62] M. Angrosino, "Doing Ethnographic and Observational Research," *Doing Ethnogr. Obs. Res.*, 2011, doi: 10.4135/9781849208932.
[63] J. Kitzinger, "Qualitative Research: Introducing focus groups," *Bmj*, vol. 311, no. 7000, p. 299, 1995, doi: 10.1136/bmj.311.7000.299.
[64] W. S. Fox and N. K. Denzin, "The Research Act: A Theoretical Introduction to Sociological Methods," *Contemp. Sociol.*, vol. 8, no. 5, p. 750, 1979, doi: 10.2307/2065439.
[65] J. D. Todd, "Mixing qualitative and quantitative methods: Triangulation in action," *Adm. Sci. Q.*, vol. 24, no. 4, pp. 602–611, 1979, [Online]. Available: http://www.jstor.org/stable/10.2307/2392366

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[66] WMA, "World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects.," *Nurs. Ethics*, vol. 9, no. 1, pp. 105–109, 2002, doi: 10.1002/9780470572757.app3.
[67] A. V. dem B. P Voigt, "The EU General Data Protection Regulation (GDPR)," *EU Gen. Data Prot. Regul.*, 2020, doi: 10.1093/oso/9780198826491.001.0001.
[68] R. Hopkins, "Information retrieval: a health and biomedical perspective," *Heal. Inf. Libr. J.*, vol. 21, no. 4, pp. 277–278, 2004, doi: 10.1111/j.1471-1842.2004.00530.x.
[69] S. Majumder and M. J. Deen, "Smartphone sensors for health monitoring and diagnosis," *Sensors (Switzerland)*, vol. 19, no. 9, 2019, doi: 10.3390/s19092164.
[70] Å. Lernmark and F. Pociot, "Genetic risk factors for type 1 diabetes," *Lancet*, vol. 387, pp. 2331–2339, 2016.
[71] S. García, J. Luengo, and F. Herrera, "Data Preprocessing in Data Mining," *Intell. Syst. Ref. Libr.*, vol. 72, 2015.
[72] E. Isikli, L. Temizer, A. E. Kazdaloglu, and E. Ari, "Time Series Analysis," *Springer Ser. Adv. Manuf.*, pp. 113–125, 2022, doi: 10.1007/978-3-030-93823-9_4.
[73] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.
[74] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2, pp. 1137–1143, 1995.
[75] J. C. TL Beauchamp, *Principles of Biomedical Ethics.*, 7th ed. 2001. doi: 10.7326/0003-4819-114-9-827_5.
[76] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Comput. Surv.*, vol. 56, no. 6, 2021, [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3457607
[77] R. C, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nat. Mach. Intell.*, vol. 1, p. 206, 2019.
[78] D. S. Char, N. H. Shah, and D. Magnus, "Implementing Machine Learning in Health Care — Addressing Ethical Challenges," *N. Engl. J. Med.*, vol. 378, no. 11, pp. 981–983, 2018, doi: 10.1056/nejmp1714229.
[79] J. Kulynych and D. Korn, "The Effect of the New Federal Medical-Privacy Rule on Research," *N. Engl. J. Med.*, vol. 346, no. 3, pp. 201–204, 2002, doi: 10.1056/nejm200201173460312.
[80] R. W. and R. V., "Big data analytics in healthcare: Promise and potential," *Heal. Inf. Sci. Syst.*, vol. 2, no. 1, 2014.
[81] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," *MCC'12 - Proc. 1st ACM Mob. Cloud Comput. Work.*, pp. 13–15, 2012, doi: 10.1145/2342509.2342513