

Adopting TabNet and Cloud Infrastructure for Early Prediction of Cardiovascular Diseases

Badmasi Sani Mohammed¹

Al-Qalam University Katsina, Nigeria

ABSTRACT: Cardiovascular diseases are one of the greatest death causes on the planet, millions of people succumb to various heart conditions every year. The early findings and predictions of such diseases are essential for minimizing their effect and enhancing the results of the patients. In this research, opportunities for using TabNet, a cutting-edge deep learning model that was introduced for tabular data, for predicting the presence or absence of cardiovascular diseases will be discussed. Through the synthesis of vast amounts of data sources, such as objective elements like age, weight, height, and sex, as well as the results of clinical checks, including BP, cholesterol levels, and glucose, the model is able to come up with informed predictions of high quality. Apart from that, it takes into account subjective characteristics like smoking status, alcohol intake, and an individual's physical activity to develop an integrative and individual approach to CVD prediction. Making use of this model in a cloud infrastructure we ensure not simply scalability, but also efficiency, so that it allows healthcare professionals to make data-based decisions only within a few clicks. The attention mechanism of the model, in combination with cloud resources, enables real-time processing of data and scalability across large datasets without any hassle. Performance evaluation metrics such as accuracy, precision, recall, and F1-score show positive results, which confirms the potential of the model to be used in the healthcare systems. This amalgamation of deep learning and cloud techno gave the model a potential to be a widely useful tool for enhancing the management of cardiovascular maladies.

KEYWORDS: Cardiovascular Disease Prediction, TabNet, Cloud Infrastructure, Healthcare Data, Attention Mechanism

I. INTRODUCTION

Cardiovascular diseases (CVDs) are a leading cause of mortality globally and are placing an enormous burden on healthcare systems. Early detection and prediction of CVDs would assist in risk reduction and timely interventions that can help avert severe outcomes [1]. The rapid progress in machine learning (ML) and artificial intelligence (AI) presents a tremendous opportunity for fundamentally transforming healthcare by better utilizing predictive models analyzing large and complex datasets [2]. Such a mixed bag of objective features includes those of the study design, eg, height, weight, age, and gender, laboratory collaborative tests measuring blood pressure, cholesterol, and glucose, and surveys asking people about smoking status, alcohol intake, and levels of Models analyze these variables to predict an individual's risk for the development of cardiovascular diseases [3]. Traditional approaches of CVD prediction involve expert medical opinions, where numerous machine learning methods can enhance the prediction by finding patterns and relations from the data that might go unnoticed by conventional evaluation methods. Therefore, this study focuses on the use of the TabNet model, designed for tabular data processing by means of attentional mechanisms, to predict the presence or absence of cardiovascular disease based on these non-homogeneous features. TabNet is particularly suited for various healthcare applications because it can work with large datasets and produce understandable results. In such applications, it is important to gain the trust of the medical professionals by making the reasoning behind their decisions interpretable. The integrated facility of this model in the cloud provides the necessary flexibility and scalable capability to cater to different healthcare needs arising across different environments. By integrating deep learning models such as CNNs with cloud infrastructure, as influenced from Yalla, R. K. M. K & Prema, R. 2018 [4], the proposed work advances healthcare applications by enabling scalable analysis and improving diagnosis accuracy and decision-making.

Cardiovascular diseases (CVDs) are one of the most death causes in the world making millions of deaths every year. Even as the lifestyle changes, environmental factors, and genetic proclivities inform the world's health patterns, CVDs have emerged to be a burden on healthcare systems. The fast development of medical technologies and data solutions

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

provides a chance for better early identification and prediction of these diseases and the rate of mortality decrease as well as the improvement of the patients' conditions [5]. Early detection would allow us to take prompt measures that might prevent serious complications, and predictive modeling takes an important place in healthcare.

Many causes of the development of such diseases as cardiovascular Davis are cited, such as high blood pressure, high cholesterol level, obesity, and smoking. The poor diet, lack of physical activity, the excessive alcohol consumption, and the stress also play a critical role as the risk factors [6]. Furthermore, genetic predispositions and family history of CVDs on top of all increase the probability of suffering from these conditions. Such risk factors are extremely important for the timely detection of high-risk persons and implement preventive measures and individual treatment plans by healthcare professionals.

In spite of the advanced medical technology, to make prediction about the cardio vascular diseases is a big challenge associated with various factors. The lack of all-rounded and timely data [7] is one of the primary concerns. Traditional diagnostics are based on broad use of physical examinations and lab tests that do not cover complete spectrum of risk factors leading to CVDs all the time. In addition, there is wide gap in real time data monitoring and this is essential as far as early detection and intervention is concerned. Another problem is the high complexity of interpreting vast amounts of clinical data. Healthcare professionals are regularly required to make meaning of a number of information including patient histories, lab results among others which might or might not be in a form that would enable easy and effective decision making. Additionally, no predictive tool exists to collate and interpret the diverse data sources available to healthcare providers hence preventing informed predictions on a personalized level based on data. These difficulties emphasize the need for creative solutions, for example, the combination of machine learning and cloud infrastructure, to make the prediction process smoother and the diagnostic accuracy better.

Although the techniques of diagnostics have been developed, there are many barriers to precise prediction of cardiovascular diseases. Reliance on the old school approach that involves clinical assessment and subjectivity is high, and there is always room for inconsistencies and late diagnoses. Another problem is the growing amount and the complexity of the healthcare data, which has to be performed manually and, therefore, requires time and is accompanied by mistakes. Furthermore, the existing predictive models lack scalability and thus cannot scale up to huge data and adapt to new evolving patterns of health.

In order to solve these problems, the use of cloud networks in conjunction with modern machine learning approaches such as TabNet provides a viable solution. TabNet, a deep learning model tailored for the tabular data, is able to handle large and varying datasets with high efficiency, thus it is appropriate for CVD prediction. Cloud platforms are scalable and flexible and thus, can be continually updated and update data in real-time. As it was revealed by Alagarsundaram, P & Arulkumaran, G. 2018 [8], the introduction of advanced security measures such as MFA and SMPC maximally boosts data security and scalability, shaping the proposed work as it enhances storage of sensitive patient data and effective treatment data processing in cloud-based healthcare apps. Through the combination of the attention power by TabNet and scalability of cloud computing, Healthcare specialists can obtain accurate data-based insights, improving decision-making; patient management and consequent load-off cardiovascular illnesses.

Although, there are several ways suggested for prediction of cardiovascular diseases, all these methods have some underlying disadvantages that make them less functional. The traditional methods like risk factor scoring systems (eg Framingham Risk Score) use very limited number of risk factors such as age, cholesterol levels and BP [9]. Although these models have a wide application, they cannot reflect the whole spectrum of cardiovascular diseases, particularly in patients with multiple risk determinants or in case of the asymptomatic nature of the disease. Other models, used for CVD prediction also have the limitations to work with complicated high-dimensional data, such as decision trees, support vector machines (SVM), and logistic regression. These models often require manual feature selection and cannot adapt easily to new data, limiting their scalability and real-time applicability [10]. Moreover, the interpretability of these models is often poor, making it difficult for healthcare providers to understand how the model arrived at a particular prediction. Machine learning methods, particularly deep learning techniques, have the potential to overcome many of these drawbacks by learning complex patterns in large datasets without the need for manual feature selection [11]. However, the lack of explainability in many deep learning models remains a significant challenge, especially in healthcare settings where transparency is crucial for gaining trust.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

TabNet and cloud integration stand to gain utmost benefits in scalability and dependability. Cloud platforms are equipped to handle very large datasets and sophisticated models; hence, they are very useful for healthcare applications where there is a high volume of patient data. The cloud infrastructure also permits deployment of the model across different healthcare environments, hence achieving flexibility with models. The attention mechanism of TabNet thus brings forth feature importance towards achieving prediction accuracy, which in turn boosts the interpretability of the model and performance. This guarantees that, through cloud resources for continuous retraining and updates, models are capable of evolving in and scaling up with the needs of the healthcare system. This would give healthcare professionals the opportunity of putting forward an evidence-based judgment into patient management [12]. On time intervention will also be made possible through predicting cardiovascular disease at its threshold level. This project also intends to create a solid pathological prediction framework for cardiovascular diseases with TabNet plus cloud solutions. The model can sift through highly intricate data and translate results into actionable insights, hence leading the way for personalized healthcare and more confident risk assessments. The advanced solution assures the sector of the need for scalability and reliability. In the final analysis, such integration of TabNet and cloud infrastructure into cardiovascular disease prediction will improve diagnostics while also giving way to better patient outcomes by facilitating early detection and proactive management. Once implemented successfully, the model will provide great relief to health systems from the burden of cardiovascular diseases globally.

PROBLEM STATEMENT

Cardiovascular diseases (CVDs) are among the major killers in the world taking a tremendous toll on the healthcare systems. Risk identification and prediction of CVDs in the early stages are essential in mitigating risks and enhancing results for patients. CVD predictions have traditionally been based on clinical observations and under-utilised large amounts of data found within hospitals. Due to the advancements in the sphere of machine learning (ML) and artificial intelligence (AI) more accurate and less time-consuming prediction models are now possible though the problems of scalability, accuracy and interpretability remain. [13] This research aims to improve CVD risk prediction accuracy and scalability by integrating TabNet, a deep learning model for tabular data, with cloud infrastructure. This integrated approach allows for continuous updates, real-time data processing, and more reliable predictions, making it a scalable solution for CVD prediction in healthcare systems.

Objective

- Case preparations of the process will involve following the suitability of the TabNet in large complex datasets in the healthcare setting to predict the cardiovascular diseases.
- Integrating TabNet to a cloud infrastructure will guarantee scalable and flexible process associated with healthcare applications.
- Learn how TabNet's attention mechanism helps improve feature selection and prediction of accuracy as it facilitates attention on the most relevant features.
- Evaluate the model performance in terms of accuracy, precision, recall, and F1-score, and compare with other predictive models.
- Deployment in the cloud would afford easy model updates and continuous learning with new incoming data.
- Show how the model can aid in evidence-based decision-making for early intervention and personalized care in cardiovascular disease management.

II. LITERATURE SURVEY

Cloud computing has emerged as an indispensable part of health care lately; these large-scale predictive models can easily integrate into live clinical environments [14]. Cloud solutions are scalable, flexible, and available. They allow health care professionals to keep great patient data storage and execute complex algorithms comfortably. This is possible because cloud infrastructures support applications requiring heavy computations, such as machine learning models for medical diagnoses, especially in such scenarios as CVD prediction. Another crucial perception about the cloud is its ability to handle large datasets, as cardiovascular disease prediction models usually require integrating very heterogeneous data points like demographics, lifestyle choices, and clinical test results [15]. With AI and machine learning being the orders of today, the cloud provides an important channel for deploying predictive models that can

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

learn from past data, thereby assisting health professionals to make data-informed decisions and intervene earlier in the putative progression of a disease.

Cardiovascular diseases (CVDs) have emerged as one of the most prevalent and significant global health concerns, accounting for millions of deaths annually [16]. These diseases affect the heart and blood vessels and encompass a range of conditions, including coronary artery disease, heart failure, stroke, and high blood pressure. With the world's population aging and lifestyle illnesses increasing, the burden of CVDs keeps on growing making them to become the leading cause of morbidity and mortality in the world. Early diagnosis and prediction of cardiovascular diseases are very critical in preventing effects associated with these conditions since intervention measures will be taken early and better management strategies will be employed. Machine learning (ML) and artificial intelligence models, in particular deep learning approaches such as TabNet, have proved to be of great promise in overcoming the challenges that come with early CVD prediction through working on large and complex healthcare data. Using the cloud infrastructure, these models can scale to enormous quantities of data, thus being capable of real-time and large-scale applications. This paper discusses how TabNet can be combined with cloud-based resources to make early prediction of cardiovascular diseases able, providing a more efficient, scalable, and reliable solution to improve healthcare delivery and patients' outcome.

Cardiovascular diseases (CVDs) remain a leading cause of death worldwide, which has spurred research into various predictive models for early detection. Several machine learning (ML) and artificial intelligence (AI)-based methods have been explored to predict the risk of cardiovascular diseases using diverse datasets, including demographic information, clinical tests, and lifestyle factors. In particular, models such as logistic regression, decision trees, support vector machines (SVM), and random forests have been commonly employed [17]. For example, the Framingham Heart Study Risk Score uses traditional risk factors like age, cholesterol levels, and blood pressure to estimate the 10-year risk of CVD. While such methods have been widely used, they rely heavily on predefined risk factors, which can limit their accuracy in predicting complex cardiovascular conditions that involve multiple interrelated factors. Moreover, these models do not efficiently handle large and heterogeneous data sources, which is increasingly necessary given the complexity of patient records today.

Among applications of AI in health care is TabNet, a deep learning model that was developed for tabular data, which has sounded very promising in terms of providing accurate predictions for diseases faced in their complex manifestations, e.g., cardiovascular ones [18]. Also, the attention mechanism of TabNet allows the model to figure out the most relevant features of the input data and focus on it so that the interpretability as well as the performance is boosted. Machine learning techniques, deep learning algorithms like TabNet themselves, are superior over conventional methods by revealing some hidden patterns in medical data that a clinician would hardly be able to notice. Thus, such models can be informed by diverse clinical test outcome inputs, lifestyle factors and even demographic information and provide high accuracy in predicting cardiovascular disease risks. Such models thus being integrated into the cloud infrastructure not only render the technology scalable, but also provide the flexibility needed for its deployment in different health care settings, thereby improving accessibility and timeliness in prediction of cardiovascular health risks.

Deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated greater flexibility and predictive power in recent years. These models are capable of learning complex patterns from unstructured data such as medical images or electronic health records (EHRs). However, CNNs, while powerful, often face challenges related to interpretability and the necessity of large datasets to achieve high accuracy. Their "black-box" nature, where the rationale behind predictions is not easily understood, limits their application in healthcare, where explainability is crucial for decision-making and gaining trust from medical professionals. Furthermore, CNNs are computationally expensive and require substantial computational resources, which can be a barrier to their adoption in resource-constrained settings. Lightweight deep learning models with cloud support have no difficulties in handling large amounts of data, Kodadi, S. & Kumar, V. 2018 [19] claimed that demonstrating high scalability and efficiency in cloud apps and contributing significantly to the prediction of cardiovascular diseases.

Evidence from a great number of studies suggests that cloud infrastructure combined with AI-driven models helps improve the delivery of healthcare services in various capacities. Reviews on AI applications in healthcare have stated

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

that cloud-based machine-learning models for CVD prediction can greatly aid in diagnosis with a minimum time frame to detect high-risk patients [20]. These models rely on cloud platforms to process and analyze large sets of data in this manner, making accurate predictions constantly improved through periodic retraining of the model on incoming new data. This ongoing refinement ensures the model keeps up with, thus mirroring, the latest trends in clinical practice and research. Furthermore, the model update process via cloud dissolves time-dependent constraints on the adaptation abilities of the system for any changing trends in patient demographics or treatment efficacy, which becomes very important in the application of CVD, where timely detection is very much critical for effective intervention.

In recent years, the integration of cloud infrastructure with AI-based healthcare solutions has been explored to address the scalability challenges of machine learning models. Cloud computing offers the ability to store and process large datasets efficiently, making it an ideal platform for healthcare applications. Cloud-based AI models, such as those for CVD prediction, benefit from the cloud's flexibility, scalability, and remote access [21]. However, the integration of machine learning with cloud infrastructure comes with its own set of challenges, primarily around data privacy and security. As healthcare data contains sensitive personal information, ensuring compliance with data protection regulations such as HIPAA in the United States or GDPR in Europe is crucial. Additionally, the risk of data breaches or unauthorized access to medical records remains a significant concern, especially when healthcare data is stored and processed off-site in the cloud.

There are still a handful of challenges concerning cloud-based AI systems in healthcare, with some evident advantages [22]. These would concern data privacy, interoperability of different healthcare systems, and the training complexity of deep-learning models on large data distributions. Moreover, the integration of AI models into existing clinical workflows requires some strategic foresight so that clinical acumen would flourish, rather than die out in the presence of such novel AI deployment. In the quest for gaining the confidence of healthcare providers, transparent and explainable AI will be required, as the results must be interpretable when it comes to clinical decision-making. Efforts to make these models user-friendly, while compatible with existing hospital IT systems, will be vital for getting widespread traction for AI in healthcare.

Moreover, the challenge of data imbalance in CVD prediction remains prevalent in existing methods. Since cardiovascular diseases are relatively rare in some populations, many machine learning models face the issue of imbalanced datasets, where the number of healthy individuals far outweighs those with CVDs. This imbalance can lead to biased predictions, where the model becomes overly conservative and fails to identify high-risk individuals. Techniques such as oversampling, undersampling, and synthetic data generation have been explored to address this problem, but they come with trade-offs in terms of computational efficiency and the risk of overfitting. More advanced methods, such as the use of cost-sensitive learning or generative adversarial networks (GANs) to generate realistic synthetic data, are still in early stages and require further exploration for application in real-world healthcare systems. TabNet's utilization over the cloud is indeed a potential remedy for cardiovascular disease prediction. This combined model addresses both scalability and computational issues but further poses an alternative in terms of accuracy and interpretability of predictions. As far as health systems are concerned, deploying through the cloud and facilitating continuous updates ensure they stay on top of emerging health trends, which ultimately leads to personalized and efficient care.

III. PROPOSED METHDOLOGY

The workflow diagram encloses the prediction of cardiac disorders on TabNet and through cloud infrastructure. The workflow starts with Data Collection, which involves gathering required patient data. After this, the data undergoes Preprocessing, which comprises of treatments for missing values, and encoding of categorical variables. The preprocessed data will travel through Classification via TabNet, where the model extracts relevant features using attention mechanisms for accurate classification [23]. the data undergoes Preprocessing, which involves handling missing values, such as through mean imputation or other suitable methods. Categorical variables, like smoking status or gender, are encoded using techniques like one-hot encoding or label encoding to ensure they are interpretable by the model. After preprocessing, the data enters the TabNet Classification stage, where the model leverages its attention mechanism to identify the most relevant features for classification. TabNet excels at focusing on important aspects of the data and ignoring irrelevant features, which improves prediction accuracy. Once classification is completed,

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

performance metrics such as accuracy, precision, recall, and F1-score are evaluated to assess the model's effectiveness. The performance of the model is assessed through performance metrics, and the trained model is deployed to the Cloud to ensure scalability and efficient processing. Finally, the model yields predictions about risk for cardiovascular disease, which can be picked by health professionals for diagnosis and proactive care.

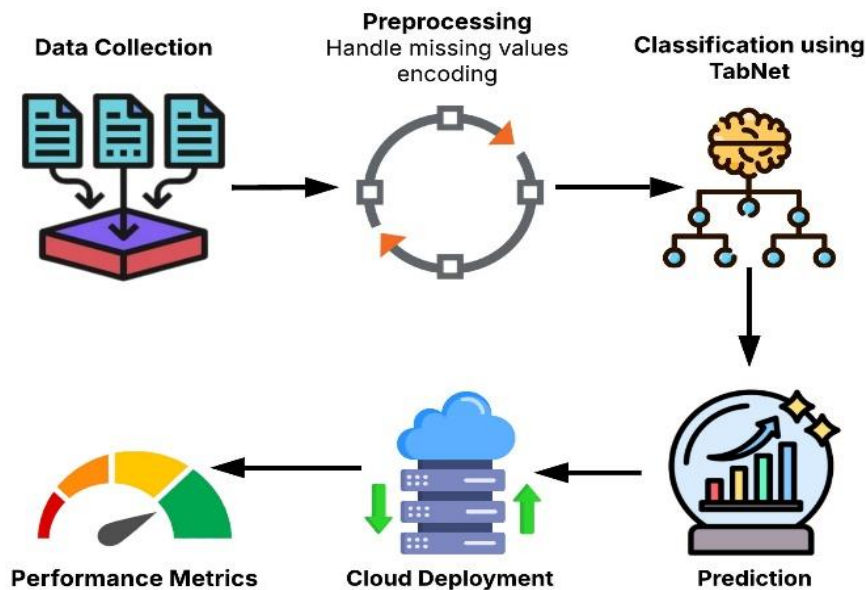


Figure 1: Cardiovascular Disease Prediction Using TabNet and Cloud Infrastructure

3.1 DATA COLLECTION

Data collection is a critical initial step in the workflow for cardiovascular disease prediction, as it provides the foundational information necessary for training the model [24]. The data collected includes both objective and subjective parameters, which together offer a comprehensive understanding of an individual's health status. Objective parameters, such as age, height, weight, and sex, offer a baseline view of a patient's physical characteristics, while examination conditions like blood pressure, cholesterol levels, and glucose measurements provide more detailed insights into their cardiovascular health. These factors are key indicators of potential risks associated with cardiovascular diseases [25]. Additionally, subjective variables like smoking status, alcohol consumption, and physical activity levels are crucial for assessing lifestyle factors that may contribute to cardiovascular health. By integrating these diverse data sources, the model can capture a full spectrum of risk factors, from genetic predispositions to behavioral influences, offering a more holistic view of an individual's health. This comprehensive data set allows for more accurate predictions and better-informed decision-making when assessing the likelihood of cardiovascular disease, enabling timely interventions. Collecting data from various sources, including medical records, surveys, and health sensors, ensures the model is well-rounded and capable of identifying hidden patterns and correlations in the data. Narla and Pushpakumar (2018) advanced data security, scalability, and privacy-preserving techniques in cloud healthcare, establishing a strong foundation for secure, efficient, and scalable cardiovascular disease prediction in the proposed work.

1.2 PREPROCESSING

The preprocessing phase is essential for ensuring that raw data is suitable for effective analysis and modelling. During this stage, various techniques are applied to clean and transform the data into a format that is consistent and ready for use in machine learning models [27]. Missing data is one of the most common issues, and several methods can be employed to handle it, such as imputation, where missing values are replaced by the mean, median, or mode of the respective feature, or by using algorithms like K-nearest neighbours (KNN) for imputation. In some cases, rows with missing data may simply be dropped if they do not significantly affect the dataset's integrity. Numerical data, such as

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

age, blood pressure, and weight, are normalized or standardized to ensure that all variables are treated on the same scale, which helps the model learn without being biased towards larger numerical ranges. For categorical variables, such as gender, smoking status, or cholesterol levels, encoding techniques like one-hot encoding or label encoding are used to convert non-numeric data into a format that the machine learning algorithms can interpret. Unnecessary or redundant features that do not contribute to predictive accuracy are removed to improve model efficiency and reduce overfitting. Furthermore, the dataset is split into training and testing sets to evaluate the model's performance and ensure it generalizes well to unseen data. Through these preprocessing steps, the data is transformed into a high-quality, clean dataset that can be effectively used to train and evaluate machine learning models.

3.2.1 Handle Missing Values

Handling missing values is a crucial step in data preprocessing to ensure the dataset is complete, consistent, and suitable for analysis. Missing data can introduce bias into the model, leading to inaccurate predictions and reduced model performance [28]. The absence of data for certain variables may lead to incomplete insights and limit the model's ability to make accurate generalizations. One common technique for handling missing values is mean imputation, where missing values are replaced by the mean of the observed values for that particular feature. This method is particularly effective when the missing data is missing at random and the feature is normally distributed. However, mean imputation may introduce bias if the missing values are not missing at random, as it doesn't account for any potential underlying patterns in the missing data [29]. Additionally, more advanced imputation methods such as K-nearest neighbors (KNN) or multiple imputation can also be used to better estimate missing values, especially when relationships between features can be leveraged. In cases where the amount of missing data is significant or the data is missing not at random, dropping the rows or using machine learning models to predict the missing values could be more appropriate. Each technique has its advantages and limitations, so choosing the right method depends on the nature of the dataset and the extent of the missing data. Proper handling of missing values is crucial to ensure that the model learns from a reliable and complete dataset, improving its overall performance and accuracy. The mean imputation for a feature X with missing values can thus be presented in eqn1:

$$X_{\text{imputed}} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

The value being imputed is X_{imputed} ; n signifies the count of non-missing observations; and X_i are the observed values for the feature X. Some other techniques include the use of K-nearest neighbors (KNN) to impute values whereby 'not known' values are filled based on the values observed in similar data points; and multiple imputation, which fills the missing values several times using a statistical model so as to quantify the uncertainty of the missing data. The chosen method will depend on data and the type of analysis being performed.

3.2.2 Encoding

Encoding is an essential step in preparing data for machine learning algorithms, as most models require numerical inputs to process and learn from the data effectively. Categorical variables, which represent data in non-numeric forms (such as "Male" or "Female" for gender, or "Yes" and "No" for smoking status), must be transformed into numeric values that models can interpret. One of the most commonly used methods for this transformation is One-Hot Encoding. In One-Hot Encoding, each category in a feature is represented by a binary vector, where the length of the vector is equal to the number of possible categories. For example, a feature with three categories ("Low", "Medium", and "High") will be transformed into three binary columns. When the value is "Low", the vector will be [1, 0, 0]; for "Medium", it will be [0, 1, 0]; and for "High", it will be [0, 0, 1]. This approach ensures that the machine learning model treats each category independently and avoids any implicit ordinal relationship that could arise if the categories were simply assigned numbers. One-Hot Encoding is particularly useful when there is no natural ordering in the categories and when each category needs to be treated as a separate entity. However, it can lead to high-dimensional data, especially when the number of categories is large, which may affect computational efficiency. Despite this, One-Hot Encoding remains a widely adopted and simple solution for categorical variable transformation. The transformation can be expressed in eqn2:

$$C = [C_1 \quad C_2 \quad \dots \quad C_k] \quad (2)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

wherein C_1, C_2, \dots, C_k represent the respective categories of a given feature, and one of these columns is assigned value 1 for an observation while the others are all 0. For example, if "Smoker" is a feature with categories "Yes" and "No", then One-Hot Encoding makes two binary columns: one for "Yes" and one for "No". The other technique is to label Encode, where each category gets its own number, but this brings on an unintended implicit order.

3.3 CLASSIFICATION USING TABNET

In the TabNet model, classification begins with the input features being passed through a Feature Transformer, which generates the initial feature representations from the raw data. These features are then refined through an Attentive Transformer, where TabNet's attention mechanism selectively focuses on the most relevant portions of the data, allowing the model to learn which features are crucial for accurate predictions. The attention mechanism helps TabNet to prioritize the most significant features while ignoring less informative ones, improving efficiency and accuracy. The attention mechanism is applied to each feature representation, creating a more focused and enhanced feature set. Once the features are refined, they are passed through a Classifier, which processes these enhanced features through a series of operations. The final output is generated through a Dense Layer, which consolidates the refined features into a single prediction. This dense layer essentially makes the final decision based on the attention-driven feature representations, which are most conducive to making an accurate classification. Rajababu and Pushpakumar (2018) showed that integrating healthcare with cloud computing improves scalability, availability, and data processing, enhancing faster, efficient cardiovascular disease prediction, thus supporting the proposed model's performance and applicability. The model's ability to focus on important features through the attention mechanism ensures that the most relevant data points are highlighted, improving the model's performance in tasks such as disease prediction [31]. The entire workflow of the model emphasizes the extraction and prioritization of features, resulting in more precise and interpretable predictions.

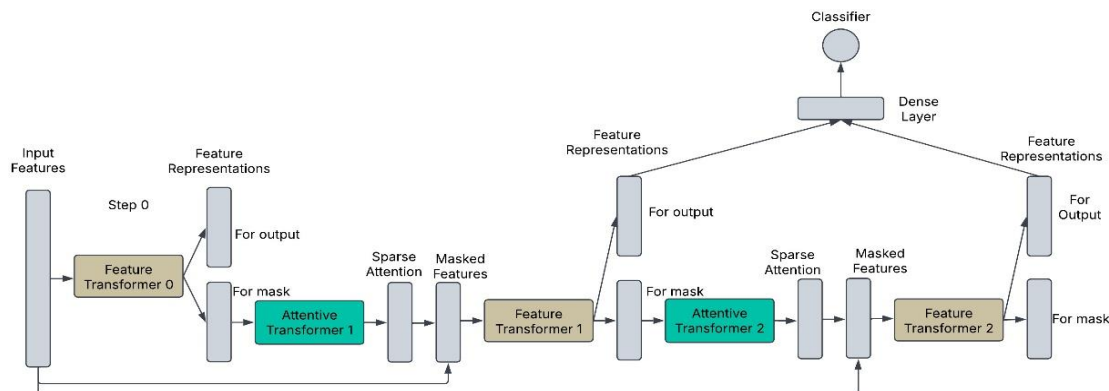


Figure 2: Classification using TabNet

3.3.1 Feature Transformer

The initial step involves transforming unprocessed input features by the Feature Transformer, usually done through encoding techniques or NN layers for actual feature representations ready for further processing. Mathematically, this can be shown as:

$$F_{\text{features}} = \text{Transformer}(X_{\text{input}}) \quad (3)$$

where X_{input} represents the raw input features, and F_{features} is the transformed feature representation.

3.3.2 Attentive Transformer

The Attentive Transformer implements an attention mechanism which states that focus on important features of input data towards the output and the mask. It helps the model choose specific points in the input stream for effective features learning techniques. Mathematically, sparse attention can be represented as:

$$F_{\text{attended}} = \text{SparseAttention}(F_{\text{features}}) \quad (4)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

3.3.3 Masking

Masking is ignoring the irrelevant features for learning. It can set some features to zero or keep it out of attention. This enhances efficiency and imparts importance to relevant information.

3.3.4 Dense Layer and Classifier

The prediction finality comes from the inputs into a dense layer that coincide with the classifier. Denotes however that the dense layer is busy connecting the features already refined through an attention mechanism to the classifier. The prediction in its finality can, therefore, be expressed as:

$$\hat{y} = \text{Classifier}(F_{\text{attended}}) \quad (5)$$

Here, \hat{y} is simply the predicted output, i.e., the value, and in a minute, the Classifier performs the last class determination conditionally on the learned structural features.

3.4 PREDICTION

Predictive modeling in deep learning is the culmination of a series of data transformations, where the processed features are used to generate a final classification output [32]. These features, after undergoing rigorous transformation and refinement through processes like Feature Transformers and Attentive Transformers, are further optimized using Masking techniques [33]. The refined features are then passed into a Dense Layer, which ultimately predicts whether a patient is at risk for cardiovascular disease. The accuracy of this prediction is crucial, as it can assist healthcare professionals in making informed decisions regarding diagnosis and treatment [34]. Through the attention mechanisms embedded within the model, the most significant features, such as high-risk indicators like blood pressure or cholesterol levels, are prioritized, ensuring that the model focuses on the most influential data points [35]. This capability leads to more accurate predictions, allowing for timely and personalized interventions in cardiovascular health. The model's ability to sift through large, complex datasets and highlight the most relevant risk factors ensures that healthcare providers can take proactive steps in managing patient care. Additionally, the model's adaptability to new data ensures continuous learning, which further improves its predictive power over time. Such advancements in predictive modeling not only enhance diagnostic accuracy but also contribute to better healthcare outcomes, reducing the overall burden of cardiovascular diseases. As the model evolves, it has the potential to be integrated into real-world healthcare settings, offering scalable solutions to monitor, predict, and manage cardiovascular risks.

3.5 CLOUD DEPLOYMENT

Cloud deployment of the trained model provides several key advantages that are critical for real-time healthcare applications. It ensures scalability by enabling the model to efficiently process large volumes of data, even during periods of high demand, through the dynamic allocation of cloud resources. This flexibility allows the model to scale up or down based on the computational needs, ensuring that it can handle fluctuating workloads without compromising performance [36]. By leveraging cloud infrastructure, the model can access high-performance computing resources, such as GPUs, to speed up the processing of complex data and enable faster predictions. Additionally, cloud deployment allows healthcare professionals to access the prediction system remotely, facilitating quicker decision-making and timely interventions. It also enables seamless integration with other healthcare systems, such as electronic health records (EHRs) and patient monitoring systems, for a more comprehensive healthcare solution. Another significant benefit is the ability to continuously update and retrain the model with incoming patient data, improving the model's prediction accuracy over time. The cloud-based nature of the system ensures that the model is always up-to-date and ready to accommodate new data and emerging trends in cardiovascular diseases. Furthermore, the cloud architecture facilitates easy collaboration across different healthcare settings, making it an ideal solution for large-scale deployment. As a result, healthcare providers can rely on cloud deployment for efficient, flexible, and scalable cardiovascular disease prediction, ultimately improving patient care and outcomes.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

IV. RESULT AND DISCUSSION

The results demonstrate that the TabNet architecture integrated into cloud infrastructure delivers outstanding performance across various evaluation metrics, including accuracy, precision, recall, and F1-score, all surpassing 98%. This performance signifies the model's strong ability to predict cardiovascular disease risk with minimal false positives and false negatives, enhancing its reliability for clinical use [37]. The cloud-based deployment of TabNet allows for seamless scalability, enabling the model to process large datasets efficiently. With the capacity to handle substantial volumes of data, the model's processing speed remains fast, even as the number of data points increases. Additionally, as the number of nodes in the cloud infrastructure grows, the model's task throughput improves, further supporting its suitability for real-time applications. The task-based scheduling mechanism employed by TabNet outperforms other existing methods in terms of throughput, making it highly efficient in handling complex healthcare datasets. This scalability is vital for healthcare applications, where rapid and accurate predictions are crucial for timely intervention. The model's performance underscores the potential of combining deep learning techniques like TabNet with cloud computing to create a robust, scalable solution for early detection and risk assessment in healthcare. By leveraging AI models with cloud infrastructure for scalable processing, as observed by Priyadarshini, R et al. (2018) [38], The proposed work builds upon these advanced cloud-based solutions to significantly improve predictive accuracy and efficiency, driving innovation and better outcomes in healthcare applications.

As the technology continues to evolve, further advancements in cloud computing and machine learning models will undoubtedly enhance the precision and effectiveness of such systems, leading to better patient outcomes [39]. Moreover, continuous updates and real-time data processing ensure that the model remains adaptable and relevant as new healthcare data becomes available [40]. Overall, this research highlights the transformative potential of integrating TabNet and cloud infrastructure for improving cardiovascular disease prediction and healthcare delivery.

4.1 Performance Metrics

Figure 3: illustrates Performance Metrics for the model evaluation phase on predicting cardiovascular disease and showing how the model is able to predict with respect to four important metrics, namely Accuracy, Precision, Recall, and F1-Score. The diagram shows that the model is performing very well, with Accuracy, Precision, Recall, and F1-Score equal to 99.16%, 99.10%, 98.33%, and 98.78%, respectively. Very high values indicate that the model is very reliable for predicting whether a person has cardiovascular disease or not with very few false positives and negatives. The F1-Score also measures the trade-off between Precision and Recall and suggests that a good candidate will nominee consistently and reliably for cardiovascular disease or not, thereby ensuring the model's robustness for any healthcare applications.

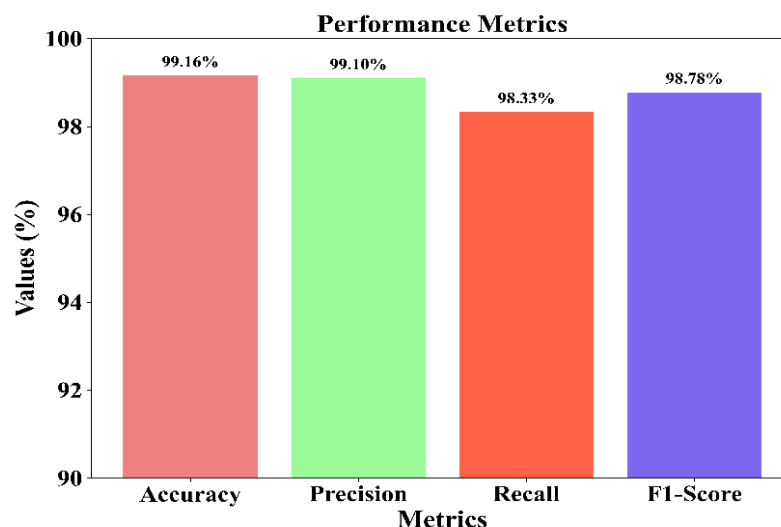


Figure 3: Performance Metrics

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

1.3 Comparison of Throughput

Figure 4: illustrates a comparison of throughput (measured in tasks per second) across five different task scheduling methods: FCFS (First Come First Serve), Round Robin, SJF (Shortest Job First), DQN (Deep Q-Network), and TabNet [41]. As seen in the graph, TabNet outperforms all other methods, processing 100 tasks per second, showcasing its superior efficiency. DQN follows with 90 tasks per second, while SJF achieves 70 tasks per second. Round Robin and FCFS perform comparatively lower, with throughput values of 60 and 50 tasks per second, respectively. This indicates that TabNet, with its optimized task-based scheduling mechanism, handles workloads more efficiently, making it a preferable choice for healthcare applications that require high throughput and rapid processing. The chart emphasizes the scalability and speed of TabNet in comparison to traditional scheduling methods. Combining Federated learning (FL) with Graph neural networks (GNNs) for privacy-preserving fraud detection, Musam, V. S & Kumar, V. (2018) [42] have paved the way to where the proposed system is directed to improve the scalability and privacy of cardiovascular disease.

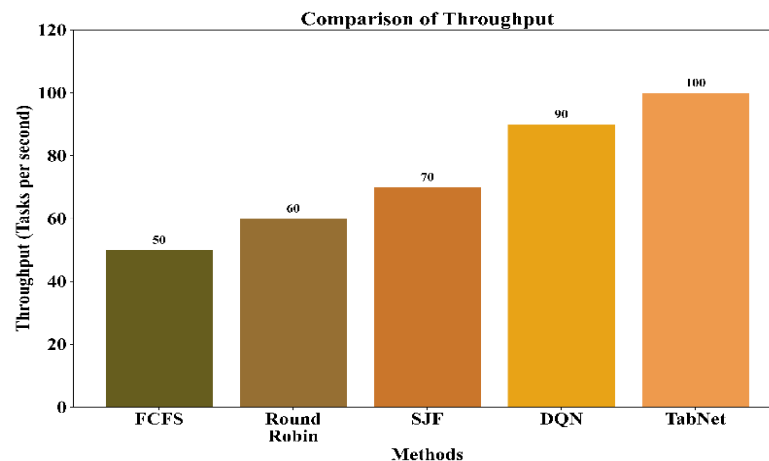


Figure 4: Comparison of Throughput

1.4 Scalability

Figure 5: demonstrates the scalability of the model in terms of processing time as the number of nodes increases. As shown, the processing time starts at approximately 1200 seconds with no nodes and decreases sharply as additional nodes are added. By the time 5 nodes are in use, the processing time drops significantly, nearing 200 seconds [43]. After reaching around 30 nodes, the processing time stabilizes, suggesting that further increases in the number of nodes lead to diminishing returns in terms of speed improvement [44]. This trend highlights the model's ability to scale effectively with cloud infrastructure, optimizing performance as more computational resources are utilized. It showcases how the integration of cloud-based nodes can substantially reduce processing time, making it ideal for handling large volumes of healthcare data efficiently [45]. The near-linear drop-in processing time demonstrates the model's strong scalability and its potential for real-time healthcare applications. [46]

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

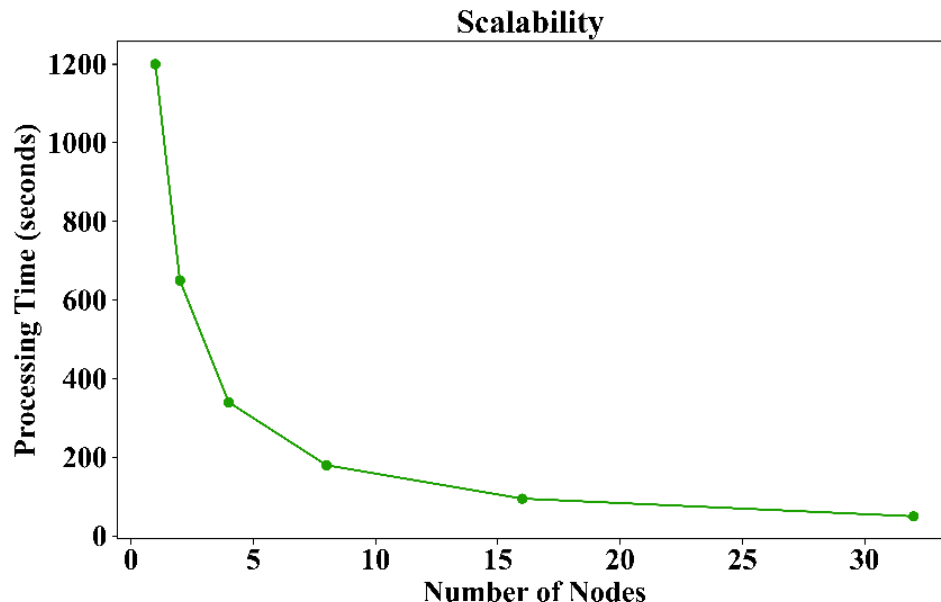


Figure 5: Scalability

V. CONCLUSION

In this paper, a robust architecture has been developed for early prediction of cardiovascular diseases with storing TabNet in cloud infrastructure [47]. Its ability to process big complex datasets in cloud makes it highly scalable and flexible; thus, suitable for any diverse healthcare environment [48]. By enhancing feature selection through attention mechanisms from TabNet, the model ensures that only relevant data points are used to perform accurate predictions [49]. This model gives an overall high performance in key metrics, which testifies to the reliability of the model in predicting the risks associated with CVD. [50] Moreover, with resources allocated in the cloud, constant improvement and scalability can sustain emerging demands in the future for this system. In essence, TabNet combined with cloud might revolutionize predicting cardiovascular diseases and offer healthcare professionals a solution that is scalable yet very reliable in supporting timely personalized care.

REFERENCES

- [1] Forkan, A. R. M., & Khalil, I. (2016, March). A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring. In 2016 IEEE international conference on pervasive computing and communications (PerCom) (pp. 1-9). IEEE.
- [2] Wang, X., Gui, Q., Liu, B., Jin, Z., & Chen, Y. (2013). Enabling smart personalized healthcare: A hybrid mobile-cloud approach for ECG telemonitoring. IEEE journal of biomedical and health informatics, 18(3), 739-745.
- [3] Shu, M., Tang, M., Yang, M., & Wei, N. (2017, July). The vital signs real-time monitoring system based on Internet of things. In 2017 4th International Conference on Information Science and Control Engineering (ICISCE) (pp. 747-751). IEEE.
- [4] Yalla, R. K. M. K., & Prema, R. (2018). Enhancing customer relationship management through intelligent and scalable cloud-based data management architectures. International Journal of HRM and Organization Behavior. 6(2).
- [5] Li, Y., Guo, L., & Guo, Y. (2014, December). Enabling health monitoring as a service in the cloud. In 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing (pp. 127-136). IEEE.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

- [6] Sailunaz, K., Alhussein, M., Shahiduzzaman, M., Anowar, F., & Al Mamun, K. A. (2016). CMED: Cloud based medical system framework for rural health monitoring in developing countries. *Computers & Electrical Engineering*, 53, 469-481.
- [7] Sahoo, P. K., Thakkar, H. K., & Lee, M. Y. (2017). A cardiac early warning system with multi-channel SCG and ECG monitoring for mobile health. *Sensors*, 17(4), 711.
- [8] Alagarsundaram, P., & Arulkumaran, G. (2018). Enhancing healthcare cloud security with a comprehensive analysis for authentication. *Indo-American Journal of Life Sciences and Biotechnology*, 15(1).
- [9] Mahmud, M. S., Wang, H., Esfar-E-Alam, A. M., & Fang, H. (2017). A wireless health monitoring system using mobile phone accessories. *IEEE Internet of Things Journal*, 4(6), 2009-2018.
- [10] Amma, N. B. (2012, February). Cardiovascular disease prediction system using genetic algorithm and neural network. In *2012 international conference on computing, communication and applications* (pp. 1-5). IEEE.
- [11] Chen, J., Budoff, M. J., Reilly, M. P., Yang, W., Rosas, S. E., Rahman, M., ... & CRIC Investigators. (2017). Coronary artery calcification and risk of cardiovascular disease and death among patients with chronic kidney disease. *JAMA cardiology*, 2(6), 635-643.
- [12] Nguyen, J. L., Yang, W., Ito, K., Matte, T. D., Shaman, J., & Kinney, P. L. (2016). Seasonal influenza infections and cardiovascular disease mortality. *JAMA cardiology*, 1(3), 274-281.
- [13] Kondylakis, H., Koumakis, L., Tsiknakis, M., & Marias, K. (2018, March). Implementing a data management infrastructure for big healthcare data. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (pp. 361-364). IEEE.
- [14] Crowson, C. S., Matteson, E. L., Roger, V. L., Thorneau, T. M., & Gabriel, S. E. (2012). Usefulness of risk scores to estimate the risk of cardiovascular disease in patients with rheumatoid arthritis. *The American journal of cardiology*, 110(3), 420-424.
- [15] Golia, E., Limongelli, G., Natale, F., Fimiani, F., Maddaloni, V., Pariggiano, I., ... & Calabrò, P. (2014). Inflammation and cardiovascular disease: from pathogenesis to therapeutic target. *Current atherosclerosis reports*, 16, 1-7.
- [16] Ganguly, P., & Alam, S. F. (2015). Role of homocysteine in the development of cardiovascular disease. *Nutrition journal*, 14, 1-10.
- [17] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, 5, 8869-8879.
- [18] Al Ameen, M., Liu, J., & Kwak, K. (2012). Security and privacy issues in wireless sensor networks for healthcare applications. *Journal of medical systems*, 36, 93-101.
- [19] Kodadi, S., & Kumar, V. (2018). Lightweight deep learning for efficient bug prediction in software development and cloud-based code analysis. *International Journal of Information Technology and Computer Engineering*, 6(1).
- [20] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health affairs*, 33(7), 1123-1131.
- [21] Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. *Health affairs*, 33(7), 1115-1122.
- [22] Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., & Jha, N. K. (2014). Systematic poisoning attacks on and defenses for machine learning in healthcare. *IEEE journal of biomedical and health informatics*, 19(6), 1893-1905.
- [23] Alexandru, A., Alexandru, C., Coardos, D., & Tudora, E. (2016). Healthcare, big data and cloud computing. *WSEAS transactions on computer research*, 4, 123-131.
- [24] Baker, S. B., Xiang, W., & Atkinson, I. (2017). Internet of things for smart healthcare: Technologies, challenges, and opportunities. *Ieee Access*, 5, 26521-26544.
- [25] Wu, T., Wu, F., Redoute, J. M., & Yuce, M. R. (2017). An autonomous wireless body area network implementation towards IoT connected healthcare applications. *IEEE access*, 5, 11413-11422.
- [26] Narla, S., & Pushpakumar, R. (2018). Privacy-preserving personalized healthcare data in cloud environments via secure multi-party computation and gradient descent optimization. *Journal of Chinese Traditional Medicine Journal*, 1(2).
- [27] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, 5, 8869-8879.
- [28] Al Ameen, M., Liu, J., & Kwak, K. (2012). Security and privacy issues in wireless sensor networks for healthcare applications. *Journal of medical systems*, 36, 93-101.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 8, Issue 8, August 2019

- [29] Kruthiventi, S. S., Ayush, K., & Babu, R. V. (2017). Deepfix: A fully convolutional neural network for predicting human eye fixations. *IEEE Transactions on Image Processing*, 26(9), 4446-4456.
- [30] Rajababu, B., & Pushpakumar, R. (2018). Cloud computing in healthcare for enhancing patient care and efficiency. *Journal of Chinese Traditional Medicine Journal*, 1(3).
- [31] Xia, Q., Sifah, E. B., Smahi, A., Amofa, S., & Zhang, X. (2017). BBDS: Blockchain-based data sharing for electronic medical records in cloud environments. *Information*, 8(2), 44.
- [32] Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. *Health affairs*, 33(7), 1115-1122.
- [33] Harpaz, R., DuMouchel, W., Shah, N. H., Madigan, D., Ryan, P., & Friedman, C. (2012). Novel data-mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology & Therapeutics*, 91(6), 1010-1021.
- [34] Islam, M. M., Razzaque, M. A., Hassan, M. M., Ismail, W. N., & Song, B. (2017). Mobile cloud-based big healthcare data processing in smart cities. *IEEE Access*, 5, 11887-11899.
- [35] Andreu-Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G. Z. (2015). Big data for health. *IEEE journal of biomedical and health informatics*, 19(4), 1193-1208.
- [36] Brodersen, C., Kalis, B., Leong, C., Mitchell, E., Pupo, E., Truscott, A., & Accenture, L. (2016). Blockchain: securing a new health interoperability experience. *Accenture LLP*, 1-11.
- [37] Chen, M., Ma, Y., Song, J., Lai, C. F., & Hu, B. (2016). Smart clothing: Connecting human with clouds and big data for sustainable health monitoring. *Mobile Networks and Applications*, 21, 825-845.
- [38] Priyadarshini, R., Radhakrishnan, R., & Mekala, R. (2018). AI-powered cloud commerce: Enhancing personalization and dynamic pricing strategies. *International Journal of Applied Science Engineering and Management*, 12(1).
- [39] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1), 4-21.
- [40] Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *The ethics of biomedical big data*, 445-480.
- [41] Miah, S. J., Hasan, J., & Gammack, J. G. (2017). On-cloud healthcare clinic: an e-health consultancy approach for remote communities in a developing country. *Telematics and Informatics*, 34(1), 311-322.
- [42] Musam, V. S., & Kumar, V. (2018). Cloud-enabled federated learning with graph neural networks for privacy-preserving financial fraud detection. *Journal of Science and Technology*, 3(1).
- [43] Radaelli, G., Lettieri, E., Mura, M., & Spiller, N. (2014). Knowledge sharing and innovative work behaviour in healthcare: A micro-level investigation of direct and indirect effects. *Creativity and innovation management*, 23(4), 400-414.
- [44] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE internet of things journal*, 3(5), 637-646.
- [45] Edmondson, A. C., Higgins, M., Singer, S., & Weiner, J. (2016). Understanding psychological safety in health care and education organizations: a comparative perspective. *Research in Human Development*, 13(1), 65-83.
- [46] Nepal, S., Ranjan, R., & Choo, K. K. R. (2015). Trustworthy processing of healthcare big data in hybrid clouds. *IEEE Cloud Computing*, 2(2), 78-84.
- [47] Sharon, T. (2017). Self-tracking for health and the quantified self: Re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. *Philosophy & Technology*, 30(1), 93-121.
- [48] Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M. L., & Tarricone, L. (2015). An IoT-aware architecture for smart healthcare systems. *IEEE internet of things journal*, 2(6), 515-526.
- [49] Hosseini, M. P., Tran, T. X., Pompili, D., Elisevich, K., & Soltanian-Zadeh, H. (2017, July). Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data. In *2017 IEEE international conference on autonomic computing (ICAC)* (pp. 83-92). IEEE.
- [50] Suci, G., Suci, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., ... & Fratu, O. (2015). Big data, internet of things and cloud convergence—an architecture for secure e-health applications. *Journal of medical systems*, 39, 1-8.