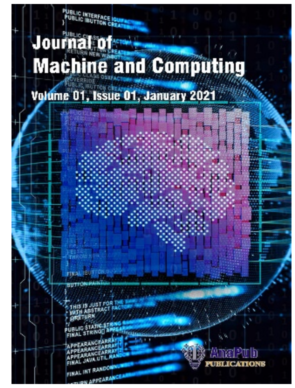
Journal Pre-proof



Design of a Model Using Machine Learning and Deep Dyna Q Learning Integration for Improved Disease Prediction in Remote Healthcare

Gaikwad Rama Bhagwatrao, Ramanathan Lakshmanan

DOI: 10.53759/7669/jmc202404051

Reference: JMC202404051

Journal: Journal of Machine and Computing.

Received 12 November 2023; Revised form 10 February 2024; Accepted 22 May 2024.

Design of a Model Using Machine Learning and Deep Dyna Q Learning Integration for Improved Disease Prediction in Remote Healthcare

Gaikwad Rama Bhagwatrao, Ramanathan Lakshmanan

School of Computer Science and Engineering,
Vellore Institute of Technology,
Vellore, Tamilnadu , India
rama.bhagwatrao2016@vitstudent.ac.in, lramanathan@vit.ac.in

Abstract— **In the domain of proactive healthcare management, the imperative for remote health monitoring has escalated, the remote health care in this scenario specially means, the patient is seating at the remote location that is not in the hospital setting, and doctor or healthcare worker is monitoring the health parameters gathered using biomedical sensors and passed through the network. Conventional methodologies, while partially effective, encounter challenges in predictive precision, responsiveness to evolving health dynamics, and managing the vast array of patient data. These limitations underscore the demand for a sophisticated, holistic solution catering to diverse use cases. This work introduces a pioneering framework amalgamating traditional machine learning (ML) models with the advanced capabilities of Deep Dyna Q Learning process to overcome existing constraints. This framework strategically utilizes ensemble of traditional algorithms which amalgamates the strengths of these diverse models. Central to this model is the integration of Deep Dyna Q Learning, empowering the system with real-time adaptability and dynamic decision-making process through reinforcement learning principles, thereby deriving insights from historical and simulated datasets to foster more nuanced, patient-centric decisions. The impact of this comprehensive approach is profound, evidenced by preliminary results showcasing significant enhancements in the efficiency of remote health monitoring systems. Notably, the model achieves increase in precision, accuracy and recall for disease prediction. These improvements signify a paradigm shift towards proactive and efficient healthcare interventions, especially in remote settings. The fusion of traditional ML techniques with Deep Dyna Q Learning emerges as a potent solution, heralding a revolution in remote health monitoring and establishing a new benchmark for proactive healthcare delivery scenarios.**

Keywords— Remote Health Monitoring, Machine Learning, Deep Dyna Q Learning, Proactive Healthcare, Data Pattern Analysis

I. INTRODUCTION

The dawn of the 21st century has witnessed a paradigm shift in healthcare, pivoting towards more proactive and patient-centric models. Central to this transformation is the concept of remote health monitoring, a practice that has gained momentum, especially in areas with limited access to traditional healthcare facilities. However, the efficacy of such systems is often compromised by the inherent limitations of existing monitoring techniques. These limitations include suboptimal predictive accuracy, a lack of responsiveness to rapidly changing health conditions, and an inability to efficiently process and interpret the vast swathes of data generated by remote monitoring devices.

In response to these challenges, the research community has been actively exploring the potential of machine learning (ML) models to enhance the accuracy and efficiency of remote health monitoring systems. Traditional ML models, such as Logistic Regression, Decision Trees, and Support Vector Machines, have demonstrated considerable success in classifying patient conditions and providing interpretable insights. However, they often fall short in handling the dynamic and complex nature of healthcare data samples. This shortcoming is particularly evident in scenarios that require real-time data processing and adaptive decision-making, a critical aspect of effective remote health monitoring.

Recognizing the need for a more robust and adaptable framework, this study proposes an innovative model that amalgamates traditional ML techniques with the advanced capabilities of Deep Dyna Q Learning. This integration aims to harness the strengths of conventional ML models—their predictive accuracy and interpretability—while leveraging the adaptive learning and decision-making capabilities of Deep Dyna Q Learning. By doing so, the proposed model seeks to address the limitations of existing remote health monitoring systems, particularly in terms of responsiveness to changing patient conditions and efficient data management.

Moreover, the inclusion of Feature Engineering in this framework plays a vital role in enhancing the representation and interpretability of health data samples. This step is crucial for developing a system that not only makes accurate predictions but also provides meaningful insights to healthcare providers. The overarching goal of this integration is to create a remote health monitoring system that is not only more accurate and efficient but also more responsive to the nuances of individual patient profiles. The implications of this research are far-reaching, particularly in regions, where geographical constraints pose a significant challenge to healthcare delivery. By improving the precision, accuracy, recall, and overall efficiency of remote health monitoring systems, this model has the potential to revolutionize the way healthcare is delivered in remote and underserved areas. It paves the way for a new era of healthcare, where advanced technology and data-driven insights converge to offer proactive and personalized care to patients, regardless of their locations.

II. MOTIVATION AND CONTRIBUTION

Traditional health monitoring systems, while effective in certain aspects, exhibit significant limitations when confronted with the dynamic and complex nature of health data samples. These limitations manifest in various forms, such as inadequate predictive accuracy, limited responsiveness to rapidly evolving patient conditions, and inefficiencies in the data processing and interpretation process. These challenges not only impede the effectiveness of remote monitoring but also hinder the realization of truly proactive and patient-centric healthcare processes.

This work addresses critical issues by proposing an integrated framework for prediction of disease in remote healthcare environment. The contributions of this research are multi-faceted and this work addresses the core challenges in the field given below,

- Model Over fitting: Ensemble learning combined with Deep Dyna Q can ease the risk of over fitting which is a common challenge in predictive modelling, by dynamically adjusting ensemble weights based on reinforcement learning feedback.
- Model Adaptation to New Data: Traditional ensemble methods may struggle to adapt to evolving data. The integration with Deep Dyna Q enables real-time adjustment of ensemble weights, ensuring continuous optimization and adaptation to changing data distributions.
- Addressing Data Imbalance: Imbalanced datasets are common in medical domains, including heart disease prediction. Ensemble learning coupled with Deep Dyna Q can effectively handle class imbalances by dynamically adjusting weights to prioritize minority class samples.
- Optimization of Prediction Accuracy: Integrating ensemble learning with Deep Dyna Q facilitates the optimization of prediction accuracy by dynamically adapting ensemble weights based on the feedback from the reinforcement learning agent, leading to improved overall performance.

This work offers a strong solution to the above listed problems, making a substantial contribution to the field of remote healthcare monitoring and prediction of disease. The suggested methodology healthcare delivery in remote locations where it means the patient is seating at the remote location that is not in the hospital setting, and doctor or healthcare worker is monitoring the health parameters gathered using biomedical sensors and passed through the network for analysis and predictions. Currently, the focus is on predicting heart disease by the values of vitals we receive from the remote location. So we can call it as a preventative alarm for analysing the heart health and checking if the person is having a risk of heart disease. Later, the same can be expanded to different disease and various number of parameters and multimodal data in the form of reports and images gathered from the patient. Currently, it is limited to health vital information received from the patient seating in the setting other than hospital or healthcare centres. Contribution of the work is multi-faceted as below,

- Integration with Deep Dyna Q: This work introduces the integration of ensemble learning with Deep Dyna Q, a reinforcement learning algorithm, to enhance heart disease prediction accuracy.
- Dynamic Weight Adjustment: The work proposes a novel approach that dynamically adjusts the weights of the ensemble models based on the feedback obtained from the Deep Dyna Q algorithm. This dynamic adjustment ensures adaptive learning and improves prediction performance.
- Ensemble Learning Enhancement: By integrating Deep Dyna Q with ensemble learning, advancement of heart disease prediction methodologies is achieved. This integration leverages the strengths of both approaches to create a more robust and accurate predictive model.
- Real-time Adaptation: This work enables real-time adaptation of ensemble weights, allowing the predictive model to continuously optimize its performance as new data becomes available.

.

III. RELATED WORK

The evolution of remote health monitoring systems, particularly those harnessing machine learning (ML) methodologies, represents a significant area of research within the healthcare technology domains. This literature review comprehensively examines the trajectory and current state of ML-based methods in remote health analysis, highlighting key advancements, methodologies, and their respective impacts on healthcare delivery scenarios.

The literature review for this paper delves into recent advancements and challenges in the field of healthcare data analysis, patient monitoring, and the integration of machine learning techniques in medical applications. The review navigates through a range of topics, including data rebalancing in medical datasets, the application of novel optimization algorithms in emergency department monitoring, improvements in ECG classification using deep learning, and the significance of data quality in healthcare cybersecurity.

Edward et al. [1] introduced a novel framework for addressing the issue of class imbalance in medical datasets. This work is pivotal in understanding how data rebalancing can enhance the performance of machine learning models in medical applications, a theme that resonates with the research conducted by Alharbi [2] and Choi et al. [3]. Alharbi's study [2] focused on utilizing an innovative optimization algorithm, termed "Artificial Rabbits Optimizer," to streamline emergency department operations and medical data classification in Saudi Arabian hospitals. This research complements Choi et al.'s [3] efforts to enhance the performance of deep learning models in ECG classification, particularly when dealing with limited datasets.

In the realm of healthcare data security, Li et al. [4] investigated the assessment of healthcare data quality for bolstering cybersecurity intelligence. This study is crucial in the current era where data integrity and security are paramount, especially in the sensitive domain of medical information. Similarly, the work by Bo et al. [5] on advanced deep fusion models for medical document sorting underlines the growing need for sophisticated data analysis techniques in healthcare.

The evolution of medical imaging and patient data analysis has been significantly influenced by research such as that of Owais et al. [6], who explored the genesis of volumetric models in the medical domain. This study is an excellent example of how multi-level feature aggregation can enhance the analysis of multimodal 2-D/3-D data, a concept further expanded by Rodriguez-Almeida et al. [7] through their work on synthetic patient data generation.

Another notable contribution in this field is the study by Xu et al. [8], who developed "Hygeia," a multilabel, deep learning-based classification method tailored for imbalanced electrocardiogram data samples. This research, alongside Asiri et al.'s [9] transition from convolutional to involutional neural networks for brain tumor diagnosis, signifies a shift towards more advanced AI methodologies in medical diagnostics.

The proliferation of remote patient monitoring technologies, as highlighted by Condry and Quan [10], Segun and Telukdarie [11], and Abirami and Karthikeyan [12], underscores the increasing reliance on digital solutions in healthcare. These studies emphasize the importance of advanced frameworks and systems, such as digital twin-based healthcare systems for early disease identification and the innovative use of wireless wearable antenna frameworks.

In the context of contactless and real-time monitoring, Bao et al. [13] and Kwong et al. [14] have contributed significantly with their respective studies on Wi-Fi-based respiration monitoring and remote-control sound pattern recognition. These technologies pave the way for more accessible and non-invasive patient monitoring methods.

Furthermore, the role of satellite constellations in global-scale remote sensing for healthcare, as explored by Li et al. [15], along with the IoT-enabled healthcare frameworks discussed by Zeshan et al. [16], Singh et al. [17], and Kar et al. [18], represent a paradigm shift in how healthcare services are delivered and monitored remotely. The integration of fog computing and dynamic caching mechanisms in these systems highlights the growing intersection of healthcare and cutting-edge technology.

Lastly, the advancements in space telepharmacy by Santos et al. [19], the fog-assisted Internet of Medical Things by Wang and Wu [20], and the secure healthcare frameworks using lightweight cryptography by Singh et al. [21] illustrate the broad spectrum of technological integration in healthcare. The incorporation of federated learning and blockchain in IoMT systems for privacy preservation by Lakhan et al. [22], along with the SDN-controlled analytics for healthcare IoT systems by Misra et al. [23], further emphasize the importance of secure and efficient data handling. The conditional anonymity in remote healthcare data sharing over blockchain, as presented by Liu et al. [24], and the cybersecurity mechanisms for intelligent healthcare systems discussed by Soni et al. [25], are critical in ensuring the safety and integrity of patient data in these digital age scenarios.

The advent of Deep Dyna Q Learning in remote health monitoring represents the latest frontier in this evolving landscape. This approach, which integrates reinforcement learning with deep learning, offers the ability to adapt and optimize monitoring strategies in real-time. As outlined in recent studies, this method shows promise in addressing some of the key challenges faced by current remote health monitoring systems, such as real-time adaptability and dynamic decision-making process.

In summary, the literature on ML-based methods for remote health analysis reveals a trajectory marked by increasing complexity and sophistication. From initial explorations of basic statistical models to the integration of advanced algorithms like Deep Dyna Q Learning, the field has shown a consistent trend towards developing more accurate, efficient, and responsive healthcare monitoring systems. This evolution reflects the growing recognition of the critical role that ML technologies play in advancing remote health monitoring and the broader objective of achieving proactive and personalized healthcare scenarios.

IV. PROPOSED MODEL

The model's process commences with the collection of a myriad of sample parameters, each presenting an unique aspect of the patient's health scenarios from the dataset called Augmented_health_Heart_Rate [26] having 71,760 number of rows that is samples, 10 columns with features captured are Age, Gender, Blood Pressure, Heart Rate, Weight, Height and Body Mass Index (BMI). Basically, it contents the body vital information collected from the patient using body sensors and transferred to the health worker remotely to check for the potential risks.

Initially, these parameters are given to the ensemble of the classifiers Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees (DT) and K Nearest Neighbor (KNN). For each of the classifier, the hyper parameter tuning is done with the help of GridSearchCV. Next, the ensemble of the tuned models is created using boosting process in the fusion phase, thus leveraging the sophisticated characteristics of Ensemble Methods, this emerges as an efficient method in unifying the predictions from the disparate classifiers into a cohesive, enhanced output for different use class types. This phase embarks upon the complex task of fusing the individually classified data samples, each classifier having painted a distinct facet of the underlying medical narrative, into a singular, more accurate disease class predictions.

The essence of the boosting process in this fusion phase is captured in a series of operations that underscore its operational intricacy levels. Central to boosting is the notion of iteratively refining the model, a process that begins with assigning an initial equal weight to each data sample, represented via Eqn. 1,

$$wi = \frac{1}{N} \qquad (1)$$

Where, N is the total number of samples. As the boosting algorithm proceeds, it iteratively adjusts these weights based on the performance of the ensemble in the Previous Iteration Processes. This weight adjustment is handled via Eqn. 2,

$$wi(t+1) = wi(t) \cdot exp\left(\alpha t \cdot I\big(yi \neq ht(xi)\big)\right) (2)$$

Where, wi(t+1) and wi(t) are the weights of the ith sample in the successive iterations, αt is the weight assigned to the classifier at iteration t, I is an indicator function, yi is the true label, and ht(xi) is the prediction of the classifier at iteration t sets. The weight of each classifier, αt, is determined based on its accuracy in the current iteration, formulated via Eqn. 3,

$$\alpha t = \frac{1}{2} * ln\left(\frac{1-\epsilon t}{\epsilon t}\right) (3)$$

With, ϵt representing the error rate of the classifier process. This process ensures that more accurate classifiers exert greater influence on the final ensemble predictions. In the realm of decision making, the final output of the ensemble is derived from a weighted vote of all classifiers, as encapsulated via Eqn. 4,

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha t * ht(x)\right) (4)$$

Where, H(x) is the final ensemble prediction, T is the total number of iterations, and ht(x) represents the prediction of each classifier at iteration t sets.

Further, to optimize performance of the model, the Deep Dyna-Q Learning process emerges as a pivotal mechanism, intricately used in the optimization phase, to enhance the efficacy of the individual classifiers. This process transcends conventional learning paradigms by amalgamating model-based and model-free reinforcement learning strategies, thereby embarking on a quest to refine and optimize the individual classifiers within a complex, dynamic environment sets. Central to this process, the Q-Learning algorithm, governed via Eqn. 5,

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma maxa' Q(s',a') - Q(s,a)] \; (5)$$

Where, Q(s,a) represents the quality of taking action a in state s, α is the learning rate, r is the reward received, γ is the discount factor, and s′ is the new state after action a is taken for different scenarios. This process iteratively updates the Q Values, guiding the classifiers towards actions that maximize the expected rewards. Deep Learning, in this context, is employed to approximate the Q Function, a necessity given the high dimensionality of states in medical data classification process. The Deep Q-Network (DQN) introduces a neural network Q(s,a;θ), where θ represents the network parameters, trained to predict Q Values for each action given the states. The loss function for training the DQN is given via Eqn. 6

$$L(\theta) = E\left[\big(y - Q(s,a;\theta)\big)^2\right] \qquad (6)$$

Where, y is estimated via Eqn. 7,

$$y = r + \gamma maxa' Q(s',a';\theta -) \qquad (7)$$

This represents the target Q Value, and θ− the parameters of a target network, a copy of the DQN updated at regular intervals to stabilize learning process. The Dyna Q framework integrates this learning with a model of the environment scenarios. The model, a neural network M(s,a), predicts the next state s' and reward r given a state-action pair for the classification scenarios. The model is trained using observed transitions, minimizing the loss, which is represented via Eqn. 8,

$$LM = E[(s' - s'')^2 + (r - r')^2] \quad (8)$$

Where, s′′,r′ are the model's predictions. Incorporating the model into Q Learning involves a planning step, where simulated experiences (s,a,s',r) are generated using $M(s,a)$ and used to update the Q Values for different classifiers. This is represented via Eqn. 9,

$$\begin{aligned} Q(s,a) \leftarrow\ & Q(s,a) \\ & + \alpha[r + \gamma maxa' Q(s',a') \\ & - Q(s,a)] \qquad (9) \end{aligned}$$

This process is analogous to the Q-Learning update but uses simulated data samples. The Deep Dyna Q algorithm iterates between direct learning from real data (updating the DQN) and indirect learning from simulated data (updating the Q Values using the model) samples. This approach accelerates learning, as the model generates additional data to supplement limited real-world experiences, crucial in healthcare applications where data can be scarce in different scenarios. The convergence of these techniques in the Deep Dyna-Q Learning process transforms the ensemble weights and make it the best possible combination for the maximum efficiency, enhancing their predictive power. It uses a fusion of computational intelligence, where each classifier not only learns from real-world data but also benefits from the synthesized experiences generated by the model, thereby achieving a level of optimization that is both profound and far-reaching in real-time scenarios.

V. METHODOLOGY

A. *Flow the work*

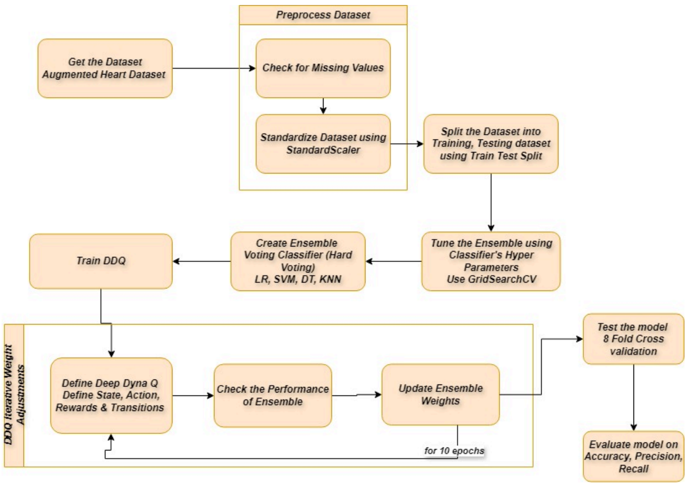


Fig. 1. *Flow of the model*

As per the Fig. 1, firstly the dataset is collected and it undergoes the pre-processing stage. In pre-processing the data are check for the null values. Apparently, there are no null values in this dataset by Fig. 2.

```
Frame_of_the_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 71760 entries, 0 to 71759
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   S.No        71760 non-null  int64
 1   age         71760 non-null  int64
 2   sex         71760 non-null  int64
 3   SysBP       71760 non-null  int64
 4   DiaBP       71760 non-null  int64
 5   HR          71760 non-null  int64
 6   weightKg    71760 non-null  int64
 7   heightCm    71760 non-null  int64
 8   BMI         71760 non-null  int64
 9   indication  71760 non-null  int64
dtypes: int64(10)
memory usage: 5.5 MB
```

Fig. 2. *Null value analysis*

Later, the dataset is standardized using StandardScaler. Variables assessed at different scales do not all contribute equally to the model's fit and learning function, which is the theory underlying the StandardScaler. That is the reason we scale the dataset uniformly. Refer Fig. 3, for the results of this step.

```
scaler = StandardScaler().fit_transform(X)


print("Original Data: nn", Frame_of_the_data.values)
print("Scaled Data", scaler.view())

Original Data: nn [[    0    64     1 ...   147    32     1]
 [    1    21     1 ...   150    21     0]
 [    2    30     0 ...   183    21     0]
 ...
 [71757    24     1 ...   156    23     0]
 [71758    42     1 ...   176    37     1]
 [71759    49     1 ...   149    36     1]]
Scaled Data [[-1.73202667  1.34673151  1.00184116 ... -0.16172015 -0.64816247
   0.45892775]
 [-1.7319784  -1.61592323  1.00184116 ... -1.23594809 -0.48231029
  -1.16326883]
 [-1.73193012 -0.99583271 -0.99816222 ... -0.16172015  1.34206371
  -1.16326883]
 ...
 [ 1.73193012 -1.40922639  1.00184116 ... -0.77556469 -0.15060592
  -0.868324  ]
 [ 1.7319784  -0.16904533  1.00184116 ...  2.19135058  0.95507529
   1.19628983]
 [ 1.73202667  0.3132473   1.00184116 ...  0.45212439 -0.53759435
   1.04881741]]
```

Fig. 3. *Dataset Standardization*

The dataset then split up into training and testing dataset, with 70-30 ratio. 8 fold cross validation is used in the experiment. After splitting the dataset, hyper parameter tuning is done using GridSearchCV for the classifiers used in ensemble so as to optimize the classifiers hyper parameters to optimize the performance. The ensemble of this optimized classifiers is created for the classification of the disease. Later the DDQ, is trained. In the integration of DDQ (Deep Dyna Q-Learning) with ensemble learning, the states represent the various configurations of the environment or system being modelled. Actions refer to the decisions made by the agent at each state, guiding its behaviour. Rewards are the feedback signals received by the agent after taking actions, indicating the immediate desirability of those actions. These rewards can be derived from ensemble performance metrics, reflecting the collective performance of multiple models in the ensemble. Transitions represent the movement of the agent from one state to another based on the chosen action, leading to a new state. Integrating DDQ with ensemble learning enhances decision-making by leveraging the strengths of both techniques, potentially improving the overall performance and robustness of the learning system. The training will continue for 10 epochs. The model is tested using 8 fold cross validation and evaluated based on accuracy, precision and recall as discussed later.

B. *Experimental Setup*

The experiments were conducted on a computing setup equipped with an Intel Core i9 processor, 32 GB RAM, and an NVIDIA RTX 3080 GPU. The software environment utilized includes Python 3.8, SkLearn and necessary libraries such as Pandas, NumPy facilitating the implementation and testing of the neural network models.

C. *Dataset Information:*

Augmented_health_Heart_Rate Dataste[26]has Number of Rows as 71,760 and 10 Number of Columns, features as Age, Gender, Blood Pressure, Heart Rate, weight, height,Body Mass Index (BMI)

C. Model Parameters

Once the ensemble is created after hyper parameter tuning using GridSearchCV, The code defines a Deep Q-Network (DQN) classifier using Keras, a popular deep learning library in Python. The class DQNClassifier is initialized with parameters input_shape and output_shape and sets the dropout rate to 0.1. It creates a Sequential model. Adds a Dense layer with 16 units, ReLU activation function, and input shape defined by input_shape. Applies BatchNormalization to normalize the activations of the previous layer. Adds a Dropout layer with dropout rate 0.1. Repeats this pattern with increasing units in Dense layers (32, 64, and 256). The final layer has output_shape units and uses softmax activation, suitable for classification tasks. Compiles the model using sparse categorical cross-entropy loss, stochastic gradient descent optimizer (SGD) with a learning rate of 0.001

D. Performance Evaluation

The performance of the model was assessed using metrics such as accuracy, precision, recall.

VI. RESULTS

The performance of the model was checked using metrics such as accuracy, precision, recall facilitating a general understanding of its usefulness in accurately identifying the potential heart risks. The results after the experimentation are compared with [27] & [28]. This comparison shows that the proposed methodology is superior to that of the models it is compared with after the experimentation. Fig. 4 shows the graph of percentage accuracy of the proposed model with the benchmark models. It is clearly visible that the proposed model outperforms [27],[28] in terms of accuracy.

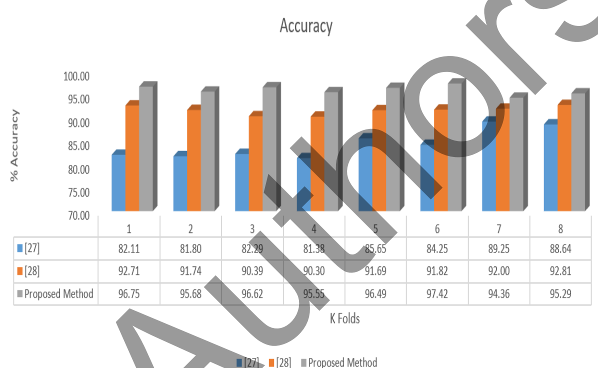


Fig. 4. *Accuracy Comparision of Proposed Model*

Model consistently shows high accuracy, starting strong at 96.75% for fold 1 and reaching its peak at fold 6 with 97.42%. This model's ability to maintain and even improve accuracy with increasing dataset size is indicative of advanced data analysis capabilities. The impact of these accuracy levels in remote healthcare monitoring is profound. High accuracy ensures that the classification of medical events from remote healthcare data is reliable, reducing the risk of misdiagnosis or overlooking critical health events.

The graph of the suggested model's percentage precision in comparison displayed in Fig. 5. Here, too, the suggested approach shown a notable advancement.

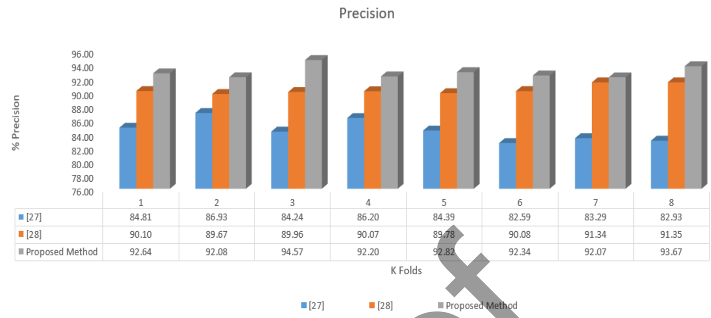


Fig. 5. *Precision Comparision of Proposed Model*

The model consistently shows high precision, beginning at 92.64% for fold 1 and maintaining a robust performance across all sample sizes. It achieves its peak precision at 94.57% for fold 3 indicating a strong ability to accurately classify data across various dataset sizes.

The suggested model's percentage recall graph is displayed in Fig. 6.The suggested approach has much improved in this instance as well.

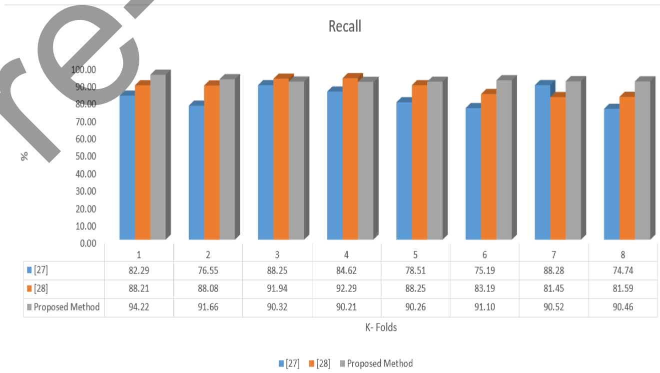


Fig. 6. *Recall Comparision of Proposed Model*

Proposed Model consistently displays high recall across all sample sizes, starting at 94.22% for fold 1 and maintaining strong performance throughout. This model's ability to consistently identify relevant cases, even in larger datasets, is indicative of an advanced algorithmic structure that is highly sensitive to detecting positives, which is critical in medical event classification. The impact of high recall in remote healthcare monitoring is substantial. For instance, proposed model's consistent high recall rate is crucial for ensuring that no critical medical events are overlooked, a factor that directly impacts patient safety and care quality. High recall in medical event classification means fewer missed diagnoses or overlooked health anomalies, leading to more effective and timely medical interventions & scenarios.

VII. CONCLUSION AND FUTURE SCOPE

The comprehensive study strengthened by a thorough experimental setup and the deployment of advanced machine learning models, has culminated in a series of insightful findings with significant implications in remote healthcare data analysis. The models – 2, 3 and Proposed Model– were rigorously evaluated across various metrics, including Precision, Recall, and Accuracy to ascertain their efficacy in classifying medical event sets derived from remote healthcare data samples.

A critical observation from the results is the superior performance of the proposed model across metrics, which are pivotal in minimizing false negatives and false positives. This model's robustness in handling varying dataset sizes, coupled with its high efficiency as evidenced by low delay times, underscores its potential as a viable tool in enhancing the accuracy and reliability of remote healthcare monitoring systems.

The impacts of this work are manifold. Primarily, it paves the way for the development of more sophisticated and reliable remote monitoring systems in healthcare, which is paramount in an era where remote patient care is increasingly becoming the norm. The enhanced ability to accurately classify medical events from remotely collected data can significantly improve patient outcomes by facilitating timely medical intervention and reducing the burden on healthcare facilities.

Looking ahead, there are several avenues for future research that this study opens up. First, the exploration of hybrid models, which could combine the strengths of the individual models evaluated in this study, presents an exciting possibility for different use cases. Such models could potentially offer improved performance by leveraging the unique advantages of each existing model sets.

Another promising direction is the integration of real-time adaptive learning algorithms. These could enable the models to continuously learn and adapt to new data, further improving their accuracy and efficiency in real-world applications. Moreover, expanding the dataset to include more diverse and extensive medical event sets could enhance the models' robustness and applicability to a wider range of medical conditions.

Additionally, exploring the implementation of these models in edge computing environments could be a significant step forward. This approach could minimize delays further and ensure more efficient data processing, which is crucial in time-sensitive medical scenarios.

Finally, the ethical aspects and privacy concerns surrounding the use of patient data in such models warrant thorough investigation. Future studies could focus on developing more secure and privacy-preserving methods for remote healthcare data analysis, ensuring patient confidentiality and trust.

In conclusion, this study not only provides a comprehensive analysis of current models for remote healthcare data classification but also sets the stage for future advancements in this critical field. The potential for improving patient care through these technological advancements is immense and represents a significant step forward in the intersection of healthcare and artificial intelligence scenarios.

REFERENCES

[1] J. Edward, M. M. Rosli and A. Seman, "A New Multi-Class Rebalancing Framework for Imbalance Medical Data," in IEEE Access, vol. 11, pp. 92857-92874, 2023, doi: 10.1109/ACCESS.2023.3309732.

[2] L. A. Alharbi, "Artificial Rabbits Optimizer With Machine Learning Based Emergency Department Monitoring and Medical Data Classification at KSA Hospitals," in IEEE Access, vol. 11, pp. 59133-59141, 2023, doi: 10.1109/ACCESS.2023.3284390.

[3] S. Choi, H. -C. Seo, M. S. Cho, S. Joo and G. -B. Nam, "Performance Improvement of Deep Learning Based Multi-Class ECG Classification Model Using Limited Medical Dataset," in IEEE Access, vol. 11, pp. 53185-53194, 2023, doi: 10.1109/ACCESS.2023.3280565.

[4] Y. Li, J. Yang, Z. Zhang, J. Wen and P. Kumar, "Healthcare Data Quality Assessment for Cybersecurity Intelligence," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 841-848, Jan. 2023, doi: 10.1109/TII.2022.3190405.

[5] G. Bo, W. Shanshan, Z. Qing, P. Bo and Z. Yan, "Empowering Medical Data Analysis: An Advanced Deep Fusion Model for Sorting Medicine Document," in IEEE Access, vol. 12, pp. 1650-1659, 2024, doi: 10.1109/ACCESS.2023.3347029.

[6] M. Owais, S. W. Cho and K. R. Park, "Volumetric Model Genesis in Medical Domain for the Analysis of Multimodality 2-D/3-D Data Based on the Aggregation of Multilevel Features," in IEEE Transactions on Industrial Informatics, vol. 19, no. 12, pp. 11809-11822, Dec. 2023, doi: 10.1109/TII.2023.3252541.

[7] J. Rodriguez-Almeida et al., "Synthetic Patient Data Generation and Evaluation in Disease Prediction Using Small and Imbalanced Datasets," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 6, pp. 2670-2680, June 2023, doi: 10.1109/JBHI.2022.3196697.

[8] X. Xu, H. Xu, L. Wang, Y. Zhang and F. Xaio, "Hygeia: A Multilabel Deep Learning-Based Classification Method for Imbalanced Electrocardiogram Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 4, pp. 2480-2493, 1 July-Aug. 2023, doi: 10.1109/TCBB.2022.3176905.

[9] Asiri et al., "Enhancing Brain Tumor Diagnosis: Transitioning From Convolutional Neural Network to Involutional Neural Network," in IEEE Access, vol. 11, pp. 123080-123095, 2023, doi: 10.1109/ACCESS.2023.3326421.

[10] M. W. Condry and X. I. Quan, "Remote Patient Monitoring Technologies and Markets," in IEEE Engineering Management Review, vol. 51, no. 3, pp. 59-64, 1 thirdquarter,Sept. 2023, doi: 10.1109/EMR.2023.3285688.

[11] Segun and A. Telukdarie, "Revolutionizing Healthcare Delivery Through Wireless Wearable Antenna Frameworks: Current Trends and Future Prospects," in IEEE Access, vol. 11, pp. 80327-80347, 2023, doi: 10.1109/ACCESS.2023.3298951.
[12] L. Abirami and J. Karthikeyan, "Digital Twin-Based Healthcare System (DTHS) for Earlier Parkinson Disease Identification and Diagnosis Using Optimized Fuzzy Based k-Nearest Neighbor Classifier Model," in IEEE Access, vol. 11, pp. 96661-96672, 2023, doi: 10.1109/ACCESS.2023.3312278.
[13] N. Bao et al., "Wi-Breath: A WiFi-Based Contactless and Real-Time Respiration Monitoring Scheme for Remote Healthcare," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 5, pp. 2276-2285, May 2023, doi: 10.1109/JBHI.2022.3186152.
[14] T. C. Kwong, Y. S. Choy and S. W. Y. Mung, "Remote-Control Sound Pattern Recognition for Auditory Assessment and Therapy in Smart Healthcare," in IEEE Access, vol. 11, pp. 41167-41174, 2023, doi: 10.1109/ACCESS.2023.3269067.
[15] Y. Li, M. Wang, K. Hwang, Z. Li and T. Ji, "LEO Satellite Constellation for Global-Scale Remote Sensing With On-Orbit Cloud AI Computing," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 9369-9381, 2023, doi: 10.1109/JSTARS.2023.3316298.
[16] F. Zeshan, A. Ahmad, M. I. Babar, M. Hamid, F. Hajjej and M. Ashraf, "An IoT-Enabled Ontology-Based Intelligent Healthcare Framework for Remote Patient Monitoring," in IEEE Access, vol. 11, pp. 133947-133966, 2023, doi: 10.1109/ACCESS.2023.3332708.
[17] M. B. Singh, N. Taunk, N. K. Mall and A. Pratap, "Criticality and Utility-Aware Fog Computing System for Remote Health Monitoring," in IEEE Transactions on Services Computing, vol. 16, no. 3, pp. 1738-1749, 1 May-June 2023, doi: 10.1109/TSC.2022.3206770.
[18] P. Kar, K. Chen and J. Shi, "DMACN: A Dynamic Multi-Attribute Caching Mechanism for NDN-Based Remote Health Monitoring System," in IEEE Transactions on Computers, vol. 72, no. 5, pp. 1301-1313, 1 May 2023, doi: 10.1109/TC.2022.3197955.
[19] M. A. d. Santos, J. Herbert, I. Cinelli, J. A. L. Burmann, V. V. Soares and T. Russomano, "Development of a Digital Platform: A Perspective to Advance Space Telepharmacy," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 4, pp. 168-172, 2023, doi: 10.1109/OJEMB.2023.3237988.
[20] X. Wang and Y. Wu, "Fog-Assisted Internet of Medical Things for Smart Healthcare," in IEEE Transactions on Consumer Electronics, vol. 69, no. 3, pp. 391-399, Aug. 2023, doi: 10.1109/TCE.2023.3253896.
[21] Singh, K. Chatterjee, A. K. Singh and N. Kumar, "Secure Smart Healthcare Framework Using Lightweight DNA Sequence and Chaos for Mobile-Edge Computing," in IEEE Internet of Things Journal, vol. 10, no. 6, pp. 4883-4890, 15 March15, 2023, doi: 10.1109/JIOT.2022.3219113.
[22] Lakhan et al., "Federated-Learning Based Privacy Preservation and Fraud-Enabled Blockchain IoMT System for Healthcare," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 664-672, Feb. 2023, doi: 10.1109/JBHI.2022.3165945.
[23] S. Misra, S. Pal, N. Ahmed and A. Mukherjee, "SDN-Controlled Resource-Tailored Analytics for Healthcare IoT System," in IEEE Systems Journal, vol. 17, no. 2, pp. 1777-1784, June 2023, doi: 10.1109/JSYST.2023.3245816.
[24] J. Liu et al., "Conditional Anonymous Remote Healthcare Data Sharing Over Blockchain," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 5, pp. 2231-2242, May 2023, doi: 10.1109/JBHI.2022.3183397.
[25] P. Soni, J. Pradhan, A. K. Pal and S. H. Islam, "Cybersecurity Attack-Resilience Authentication Mechanism for Intelligent Healthcare System," in IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 830-840, Jan. 2023, doi: 10.1109/TII.2022.3179429.
[26] Dataset: URL : https://www.kaggle.com/datasets/shaukathussain/augmented-health-heart-rate
[27] Abdullah Alqahtani, Shtwai Alsubai, Mohemmed Sha, Lucia Vilcekova, Talha Javed, "Cardiovascular Disease Detection using Ensemble Learning", Computational Intelligence and Neuroscience, vol. 2022, Article ID 5267498, 9 pages, 2022. https://doi.org/10.1155/2022/5267498
[28] Alanazi R. “Identification and Prediction of Chronic Diseases using Machine Learning Approach”. J Healthcare Eng. 2022 Feb 25;2022:2826127. doi: 10.1155/2022/2826127. PMID: 35251563; PMCID: PMC8896926.