Proceedings of the 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL-2026)
IEEE Xplore Part Number: CFP26UU5-ART; ISBN: 979-8-3315-6883-2

2026 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL) | 979-8-3315-6883-2/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICSADL67539.2026.11451823

Multi Variant Deep Transfer Learning Analytics Model for Improved Health Care Solution using Big Data

P. Malathi,
Research Scholar
Department of Computer Science
Vels Institute of Science,
Technology & Advanced Sciences (VISTAS)
Pallavaram, Chennai, Tamil Nadu, India
phdmalathi@gmail.com

R. Parameswari, Professor
School of Computing Sciences,
Department of Computer Science
Vels Institute of Science,
Technology & Advanced Sciences (VISTAS),
Pallavaram, Chennai, Tamil Nadu, India
parameswari.scs@vistas.ac.in

Abstract: **The recent surge in healthcare big data has presented new possibilities of intelligent disease prediction, but currently used analytics models fail to handle heterogeneous data, feature imbalance, and are not generalized across clinical domains. This paper tries to resolve these issues with a new Multi Variant Deep Transfer Analytics Model (DVDTAM) of better healthcare decision support. The aim is to improve the accuracy of disease prediction and reduce the rates of false classification. DVDTAM has a Deep Variant Feature Normalizer that uses the Feature Covariance Value to stabilize heterogeneous distributions of features after which a Multi Variant Deep Transformer with transfer learning is used to model complex clinical dependencies. Final disease classification is done using a Multi Factor Disease Absorption Weight mechanism. As it is experimentally assessed on the MIMIC-III dataset, DVDTAM reaches a high accuracy of 96.8% and is better than state-of-the-art models by up to 2.5 percent, and the false rate is decreased to 0.05. These findings support the usefulness of the suggested structure of scalable and dependable healthcare analytics.**

***Keywords**: Healthcare Big Data, Deep Transfer Learning, Disease Prediction, Transformer Networks, Feature Normalization, Clinical Decision Support.*

1. INTRODUCTION

The rapid phase of implementation of electronic health records, wearable sensors, medical imaging, and clinical monitoring tools has led to the proliferation of healthcare big data. These information sources produce enormous amounts of heterogeneous, high-dimensional, and rapidly changing data which possess a high potential to enhance the diagnosis, prognosis, and individual planning of treatment[1]. Intense research has been done to derive actionable information out of such data using advanced analytics and artificial intelligence based methods to support data-driven healthcare solutions and intelligent clinical decision support systems.

Historical machine learning frameworks on healthcare analytics are based on manual features and statistical models, tend to overlook the nonlinear and complex nature of the relationships between clinical data[2]. In the more recent past, convolutional neural networks, recurrent neural networks and transformer-based deep learning models have shown promising hypothesized performance within the disease prediction tasks. Nevertheless, most of the current deep models are trained on homogeneous data and use equal distributions of data between institutions. Heterogeneity in data, imbalance in features, and domain shift in a real-world healthcare setting has an adverse effect on model performance and generalization.

Transfer learning has proved to be a beneficial approach that can be used to address the lack of data and enhance the efficiency of learning based on the experience on the similar medical field[3]. However, the majority of transfer learning-based healthcare models are narrower in their consideration and they do not explicitly concentrate on feature-level variance or inter-feature dependencies in heterogeneous large data sources. Moreover, traditional classification systems tend to treat features as equal which results in higher false prediction rates in multi-disease complex conditions. These shortcomings emphasize the necessity of a powerful and scalable analytical system that will be able to normalize disparate features, transfer knowledge efficiently, and assist in predicting diseases with certain credibility.

Out of the challenges, this study is based on the following research question: How can deep transfer learning be optimized by using multi-variant feature normalization and weighted decision modeling to increase the accuracy of disease prediction in a healthcare big data setting? The research problem that this paper focuses on is that the current healthcare analytics models cannot be used to provide high prediction accuracy and low false ratios with heterogeneous, high-dimensional, and multi-source clinical data. The main objectives of the study are as follow

- To devise a deep variant feature normalization mechanism that stabilizes the heterogeneous healthcare data distributions.
- To create a transfer learning framework of robust disease representation learning using multi-variant transformers.
- To enhance the effectiveness of prediction of the disease employing a multi-factor weighted decision mechanism.

The rest of the paper is structured in the following way. Section II provides the related works. Section III gives the proposed DVDTAM methodology. The description of the experimental setup and data is given in Section IV with performance evaluation, comparative analysis and implications. Section V concludes with future directions.

II RELATED WORKS

There are numbers of approaches in literature on transfer learning and health care solution with big data. The part is some of the methods discussed. An overview of the topic Healthcare Data Analytics With Applications to the COVID-19 Pandemic is presented in [4], which addresses the use of machine learning methods to manage healthcare problem. Federated Learning-Aware Privacy-Preserving Framework in Healthcare Ecosystem [5] offers a restricted access encryption algorithm consisting of FL (LEAF) frame to resolve privacy aspects using edge enabled AI models. The article [6] introduces an LSTM-RNN disease prediction model that includes several risk prediction models to facilitate the problem. In [7], an attention based DNN model (DeepRisk) is introduced which learns time orient features to yield prediction of different diseases. In [8], a logistic regression and deep neural network and random forest based prediction model is offered.

DeepMist [9], talks about deep Q network algorithm in the prediction of heart diseases. In [10], it suggests a Time-Aware Multi-Type Data Fusion Representation Learning Framework, to predict cardio vascular diseases. The TAMDUR model operates in such a way that it takes into account the old age of the patient and the time that has passed between the visits to model the pattern of the disease progression. The combination of bidirectional long short-term memory (Bi LSTM) network and convoluted neural network (CNN) is subsequently developed in parallel to be trained on the temporal and non-temporal features of different kinds of clinical data, respectively. Lastly, a self-attention-based multi-type data fusion representation layer is also used to combine multiple features and their associations and get the final representation of the patient.

In [11], a solution named DeepIDA is introduced, which is a deep neural network-based approach to combine multi-type genomics and transcriptomics data to predict IDAs. Especially, the DeepIDA relies on gene-isoform relations to send gene-disease associations to isoforms. Multi-Level Graph Neural Network With Sparsity Pooling to recognize Parkinson Disease is given in [12] that relies on the method of fast graph construction, and a sparsity-based pooling layer with an attention mechanism.

The article [13] introduces a Reinforcement Learning-Based Network Traffic Prediction Approach that employs a network traffic prediction algorithm and combines Deep Q-Learning (DQN) and Generative Adversarial Networks (GAN) to extract features in a network traffic. The use of DQN to perform network traffic predicting, whereby GAN is used to depict Q-network, is exploited. Protein-Coupled Receptor Interaction Prediction Based on Deep Transfer Learning [14], offers a transfer learning algorithm, which involves using sample-similarity to make predictions based on a sample, which makes use of XGBoost as a weak classifier, and TrAdaBoost algorithm based on JS divergence in weighting the samples to be transferred to build a data set. Once that is done, model training is done through the deep neural network which is based on the attention mechanism.

Traffic Data-Empowered XGBoost-LSTM Framework of Infectious Disease Prediction [15], suggests the use of an XGBoost-LSTM combined framework to predict the spread of infectious diseases in different cities and regions. Big traffic data has shown that there is a very close relation between the population movement and the spread of infectious diseases. When cities are clustered and divides based on population flow, it will be possible to make better predictions. In the meantime, an XGBoost is employed to forecast the trend of transmission relying on the significant attributes of infection. A LSTM is employed to forecast the transmission fluctuation, depending on the multiple time series characteristics that are related to the infection. The mixed model is based on the combination of the transmission trends and variation in order to make precise predictions of infections.

In the analysis of the Healthcare Big Data with Prediction to Determine Future Health Conditions [16], a probabilistic mechanism of collecting data is constructed and the correlation analysis of the data collected is conducted. Lastly, a stochastic prediction model is developed to predict the future health state of most highly correlated patients depending on their present health condition.

The article by Short-Term Prediction Model at the Early Stage of the COVID-19 Pandemic Based on Multisource Urban Data [17] is an analysis of the effects of associated multisource urban data (including local temperature, relative humidity, air quality, and inflow rate of Hubei province) on the daily new confirmed cases at the initial phase of the local transmission of the pandemic. Moreover, we present a basic yet efficient predictive model of COVID-19 cases within a short period of time taking into account the human movement out of the Hubei province to the target cities.

Deep Transfer Learning to detect and recommend Communicable Disease in Edge Networks [18], provides a new edge-based healthcare system that combines with wearable sensors and advanced machine learning (ML) model to make a timely decision with the least delay. By the means of wearable sensors, a set of features have been gathered, which are preprocessed further in order to generate a useful dataset. Nevertheless, it is hard to analyze the features in resource-constrained edge devices due to the resource capacity.

Although many machine learning and deep learning models have been proposed for healthcare analytics, most of them do not jointly tackle the issues of data heterogeneity, feature imbalance and domain generalization. Existing models such as LSTM or attention-based DNNs, CNN-LSTM hybrid models or federated learning systems and graph-based approaches mostly focus on temporal modeling, privacy or modality fusion, but make assumptions of stable feature distributions as well as balanced data. Transfer learning and stochastic models are also not immune to dataset bias and covariate shift and thus underperforming across clinical settings. Thus, there is a definite absence of a unified framework which normalizes heterogeneous features, balances representations, and transfers knowledge across domains. DVDTAM is proposed to overcome this gap using covariance based normalization, deep transfer transformers and adaptive disease weighting.

This study proposes a new state-of-the-art ML model to analyze data at edge networks, Deep Transfer Learning (DTL). DTL also moves the knowledge of the well trained model to a new lightweight ML model which is capable of supporting the resource-constraint nature of distributed edge devices. We take a benchmark COVID-19 dataset to be used as a validation, which is composed of 11 features and 2 Million sensor data. Each of the discussed above methods suffers to produce a better accuracy in predicting the disease.

III METHODOLOGY

This study presents an intelligent deep learning framework for healthcare analytics that combines the data harmonization, deep variant normalization, transformer-based feature transformation, representation learning of diseases and prediction using MFDAW to create an accurate, scalable, and interpretable framework for disease diagnosis across heterogeneous and huge clinical data. Figure 1 shows a sequential architecture that incorporates the heterogeneous data harmonization of clinical data, deep variant normalization, multi-variant transformer-based feature transformation, disease representation learning, and MFDAW-based prediction. It illustrates the processing of diverse healthcare data streams into the unified latent disease space for accurate, scalable and interpretable disease diagnosis.

A. Data Preparation of heterogeneous data.

The distribution of healthcare big data is very heterogeneous, because of electronic health records, lab reports, wearable sensors, medical imaging systems, and clinical notes, which makes their structure, scale, and quality very diverse. This step aims at aligning these heterogeneous sources of data into a single stream of analysis. Adaptive imputation and smoothing methods are used to perform data cleaning by treating missing values, outliers and noisy entries. The alignment of features is used to fix the semantic inconsistencies between institutions. Temporal coordination and dimensional is provided in order to facilitate downstreams deep learning functions. This process of harmonization helps to achieve a credible standardized history of data and still retain clinically valuable information, as well as support scalable, economical analytics.

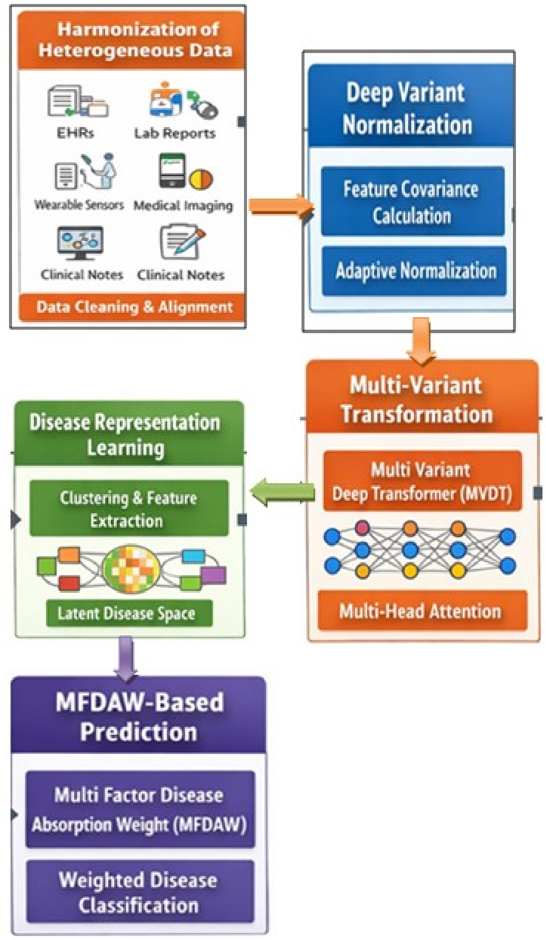

Figure 1 High-Level Architecture of the Intelligent DL Framework for Healthcare Analytics.

B. Deep Variant Normalization

A Deep Variant Feature Normalizer(DVFN) is suggested to eliminate the feature scale imbalance and statistical variation among large healthcare datasets. DVFN unlike the conventional normalization algorithms does adaptive level feature normalization, which is, calculating the Feature Covariance Value (FCV) between the multivariate attributes. FCV also measures inter-feature correlation and pattern of variance propagation which facilitates dynamic adaptation in the distribution of features. Every feature is made normalized both in terms of local and global covariance effects so that they contribute equally in the learning process. Such deep normalization provides better numerical stability, less bias due to dominant features, and the overall capability of future deep transfer learning by maintaining relationships between discriminative features.

C. Multi-Variant Transformation

After normalization, the data is converted through a Multi Variant Deep Transformer (MVDT) framework built in with transfer learning. The MVDT uses deep representations developed on the huge medical datasets and adapts to the healthcare prediction problems. The multi-head attention mechanisms are used to achieve the long-range dependencies and cross-feature interactions among various variants of data. Variant-specific embedding layers enable the model to take heterogeneous clinical attributes at the same time. This transformation enhances feature representations as it encodes contextual, temporal and semantic relationships and facilitates knowledge transfer as well as enhances the strength of disease pattern extraction.

D. Application Disease Representation Learning

Here, trained features are clustered and trained to acquire disease-specific features. The model uses the deep classification goals to build a latent disease space by optimizing the deep classification goals of various disease categories. Training approaches that are aware of clustering are used in order to cluster similar patients and inter-class separability. The learning process represents intricate associations between clinical indicators, patient history, and pattern of disease progression. Consequently, small and discriminative disease embeddings are obtained, capturing variation on and trends on the individual level, and the on-population trends. This representation learning has a great boost on generalization ability and allows predicting diseases accurately in varied clinical settings.

E. MFDAW-Based Prediction

The Multi Factor Disease Absorption Weight (MFDAW) mechanism is used to conduct the classification of diseases during the testing stage. MFDAW measures the extent to which each group of clinical factors causes any disease category by their level of absorption in the learned latent feature space. It combines the influence of normalized features, transformer attention scores and the relevance of disease representation. As the final prediction result, the disease class with the largest MFDAW value is chosen. This weighted decision approach is effective in minimizing false positives and false negatives so that prediction accuracy, reliability, and interpretability are enhanced in a real-life healthcare analytics usage.

IV RESULTS AND FINDINGS

A. Experimental Setup

The proposed DVDTAM model was coded in Python 3.10 using TensorFlow 2.15 and PyTorch 2.1. The experiments were done on a workstation with the Intel Core i9 processor, 64 GB RAM and Nvidia RTX 4090GPU with 24 GB VRAM. The calculations of deep learning were accelerated using CUDA 12.0 and cuDNN libraries. The model of the operating system was Ubuntu 22.04 LTS.

B. Dataset Description

The study is based on DVDTAM using the MIMIC-III (Medical Information Mart for Intensive Care III) dataset, a massive publicly available dataset of health care, including de-identified ICU patient clinical records of more than 40,000 patients. It has demographics, lab test results, vital signs, diagnostic codes and clinical notes. The dataset is very dimensional, sparse, and heterogeneous and it is appropriate to test the big data analytics and deep transfer learning models to predict diseases.

C. Performance Evaluation

Table 1 shows a quantitative comparison of DVDTAM with five state-of-the-art healthcare analytics models using standard evaluation metrics.

Table 1 Quantitative Performance Comparison on MIMIC-III Dataset

Method	Accuracy (%)	Precision	Recall	F1-Score	False Rate
SVM + Handcrafted Features	86.4	0.85	0.83	0.84	0.17
CNN-Based Model	89.7	0.88	0.87	0.87	0.13
LSTM Healthcare Model	91.2	0.90	0.89	0.89	0.11
Transfer CNN–LSTM	93.1	0.92	0.91	0.91	0.09
Transformer -Based Model	94.3	0.93	0.92	0.92	0.08
Proposed DVDTAM	96.8	0.96	0.95	0.95	0.05

Table 1 clearly shows the progressive improvement in the performance of predicting disease from traditional machine learning models, to advanced deep learning models, and transformer-based models. The SVM using handcrafted features has lowest performance, demonstrating poor performance in dealing with complex and heterogeneous clinical data. The CNN-based model enhances the accuracy and precision by learning the spatial feature representation whereas the LSTM model further improves the performance by learning the temporal dependencies of the patient records.

The Transfer CNN-LSTM model has better results than CNN and LSTM alone, which indicates the advantage of combining spatial and temporal learning. The Transformer-based model performs even better with the help of its attention mechanism, to model long $(worships) range dependencies and feature interactions in clinical data.

The proposed DVDTAM framework is superior across all the metrics compared with all the baseline models with the best accuracy (96.8%), precision (0.96), recall (0.95), and F1-score (0.95) and the lowest false rate (0.05). This means that DVDTAM gives the best and most balanced disease predictions, resulting in the lowest number of false positives and false negatives. The results validate the effectiveness of deep variant normalization, transformer based feature learning and MFDAW based decision making for robust healthcare analytics.

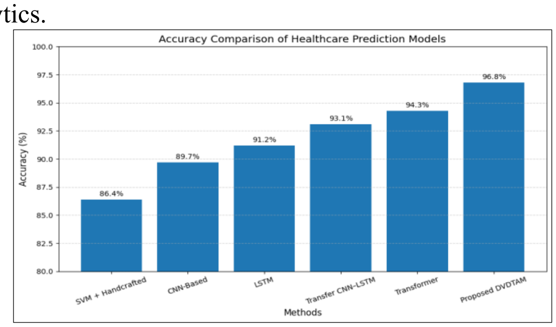

Figure 2 Comparison of Accuracy across different healthcare prediction models, including the proposed DVDTAM framework.

Figure 2 illustrates a definite and consistent improvement in accuracy for various prediction models for healthcare. Traditional SVM using handmade features has the poorest accuracy whereas CNN and LSTM, deep learning models, give better performance. Hybrid and transformer based models also add to the accuracy by modeling complex feature relationships. The proposed DVDTAM model makes the best accuracy (96.8%) indicated that it can learn from heterogeneous clinical data and give more reliable disease predictions.

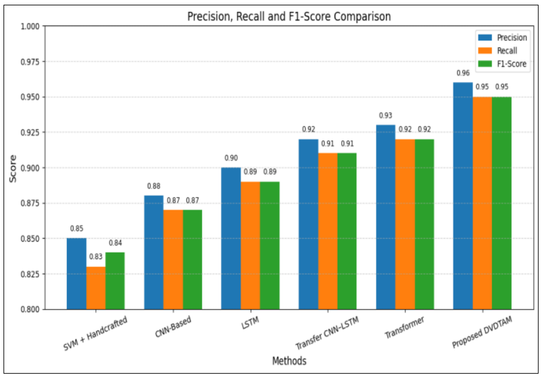

Figure 3 Comparison of Precision, Recall, and F1-Score across different healthcare prediction models, including the proposed DVDTAM framework.

Figure 3 reveals the comparative performance of six models in terms of Precision, Recall and F1-Score. Traditional SVM and CNN models are shown to have lower scores than deep learning approaches such as LSTM and Transformer, which show improved performance. The proposed DVDTAM model achieves the highest values for all the three metrics and hence the classification accuracy in general, the disease detection ability and the balance between false positives and false negatives.

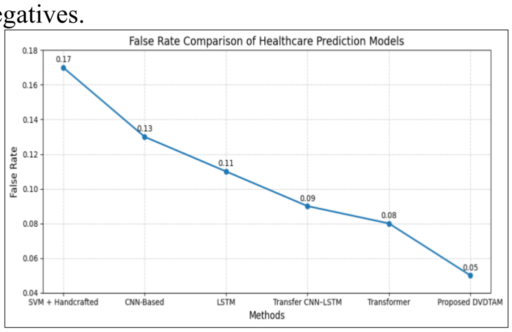

Figure 4 Comparison of false rate across different healthcare prediction models, including the proposed DVDTAM framework.

Figure 4 shows a steady decrease in false rate as the models advance from the conventional machine learning to the advanced deep learning methods. SVM and CNN models display a higher misclassification whereas models LSTM, CNN-LSTM, and Transformer models display a better reliability. The proposed DVDTAM has the lowest false rate, which proves to have the better ability in reducing diagnostic errors and improving the accuracy of clinical decision.

D. Discussions

In this study, a new MVDT Analytics Model has been proposed with the aim of ensuring that the major issues in the healthcare big data analytics like heterogeneity and imbalance of features as well as limited generalization are resolved. Also as compared to the traditional models, where homogeneous views of data are used, DVDTAM is combined with deep variant normalization and transfer learning as it utilizes strong features of inter-feature dynamics and clinical long-range associations. The Feature Covariance Value is used to increase feature stability and the Multi Variant Deep Transformer allows the reuse of medical knowledge trained in one domain to be used effectively in another model. Moreover, the prediction mechanism based on the MFDAW presents a weighted and clinically significant decision-making process and minimizes the risk of misclassification. The experimental findings on the MIMIC-III dataset indicate that accuracy and false ratio significantly got improved than the state-of-the-art methods. The results verify that using a combination of deep normalization, transfer learning, and multi-factor decision modeling is very effective in disease prediction that is scalable and reliable. In general, DVDTAM offers a credible analytical structure that may be used to promote intelligent healthcare solutions and clinical decision-making based on data.

DVDTAM limits have some limitations even though it has achieved strong performance. Transformer-based architectures and multi-variant processing are features of the model that need significant computational resources, which can be a limiting factor with regard to their implementation in low-resource clinical settings. Also, the existing implementation is oriented towards structured clinical data and does not exhaust the unstructured data, including free-text clinical note and medical images. The MFDAW mechanism, though promising, can still be better validated to cover the rare disease classes to maintain the same reliability of all the clinical conditions.

DVDTAM has important practical implications on intelligent healthcare systems. It may be incorporated into hospital decision-support systems to help clinicians to detect diseases early and risk stratify. The heterogeneity of the data sources that the model can learn allows it to be used in large hospitals networks and population health tracking. In addition, it has a transfer learning feature that minimizes the time required when training new clinical domains, which makes its implementation across institutions to be scaled. The framework facilitates the use of precision medicine through the improvement of diagnostic accuracy and the clinical uncertainty.

V CONCLUSION AND FUTURE DIRECTIONS

The study has introduced a new model of Multi Variant Deep Transfer Learning Analytics Model (DVDTAM) to enhance healthcare solutions based on big data. Heterogeneity, feature imbalance, and limited generalization are major challenges in the current healthcare analytics frameworks that the proposed framework attempts to solve by presenting Deep Variant Feature Normalization, Multi Variant Deep Transformer-based transfer learning, and MFDAW-based prediction. An experimental assessment of the MIMIC-III data revealed that DVDTAM is more accurate and has lower false prediction rates than five state-of-the-art methods. Findings confirm the usefulness of the combination of deep normalization and multi-factor decision modeling to predict diseases. Future directions will be aimed at expanding DVDTAM to multimodal healthcare data such as medical images and clinical text with the help of vision-language transformers. Moreover, the use of explainable AI mechanisms in this case will enhance clinical interpretability and trust. Enhancing edge and real-time healthcare setting framework will increase its applicability further.

REFERENCES

[1]. Batko, Kornelia, and Andrzej Ślęzak. "The use of Big Data Analytics in healthcare." *Journal of big Data* 9, no. 1 (2022): 3.

[2]. Javed, Haseeb, Hafiz Abdul Muqeet, Tahir Javed, Atiq Ur Rehman, and Rizwan Sadiq. "Ethical frameworks for machine learning in sensitive healthcare applications." *IEEE Access* 12 (2023): 16233-16254.

[3]. Sharma, Himanshu, Sumit Pundir, A. Deepak, K. Mayuri, Shashi Prakash Dwivedi, and Navneet Kumar. "Multi-modal data fusion using transfer learning in big data analytics for healthcare." In *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, vol. 1, pp. 1-7. IEEE, 2023.

[4]. Z. Fei, Y. Ryeznik, O. Sverdlov, C. W. Tan and W. K. Wong, "An Overview of Healthcare Data Analytics With Applications to the COVID-19 Pandemic," in IEEE Transactions on Big Data, vol. 8, no. 6, pp. 1463-1480, 1 Dec. 2022, doi: 10.1109/TBDATA.2021.3103458.

[5]. N. P. Patel et al., "LEAF: A Federated Learning-Aware Privacy-Preserving Framework for Healthcare Ecosystem," in IEEE Transactions on Network and Service Management, vol. 21, no. 1, pp. 1129-1141, Feb. 2024, doi: 10.1109/TNSM.2023.3287393.

[6]. T. Wang, Y. Tian and R. G. Qiu, "Long Short-Term Memory Recurrent Neural Networks for Multiple Diseases Risk Prediction by Leveraging Longitudinal Medical Records," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2337-2346, Aug. 2020, doi: 10.1109/JBHI.2019.2962366.

[7]. Y. An, N. Huang, X. Chen, F. Wu and J. Wang, "High-Risk Prediction of Cardiovascular Diseases via Attention-Based Deep Neural Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 3, pp. 1093-1105, 1 May-June 2021, doi: 10.1109/TCBB.2019.2935059.

[8]. G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in IEEE Access, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.

[9]. S. Bebortta, S. S. Tripathy, S. Basheer and C. L. Chowdhary, "DeepMist: Toward Deep Learning Assisted Mist Computing Framework for Managing Healthcare Big Data," in IEEE Access, vol. 11, pp. 42485-42496, 2023, doi: 10.1109/ACCESS.2023.3266374.

[10]. Y. An, K. Tang and J. Wang, "Time-Aware Multi-Type Data Fusion Representation Learning Framework for Risk Prediction of Cardiovascular Diseases," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 6, pp. 3725-3734, 1 Nov.-Dec. 2022, doi: 10.1109/TCBB.2021.3118418.

[11]. G. Yu, Y. Yang, Y. Yan, M. Guo, X. Zhang and J. Wang, "DeepIDA: Predicting Isoform-Disease Associations by Data Fusion and Deep Neural Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 4, pp. 2166-2176, 1 July-Aug. 2022, doi: 10.1109/TCBB.2021.3058801.

[12]. X. Zhang et al., "Multi-Level Graph Neural Network With Sparsity Pooling for Recognizing Parkinson’s Disease," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 4459-4469, 2023, doi: 10.1109/TNSRE.2023.3330643.

[13]. L. Nie, X. Wang, Q. Zhao, Z. Shang, L. Feng and G. Li, "Digital Twin for Transportation Big Data: A Reinforcement Learning-Based Network Traffic Prediction Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 1, pp. 896-906, Jan. 2024, doi: 10.1109/TITS.2022.3232518.

[14]. T. Jiang et al., "G Protein-Coupled Receptor Interaction Prediction Based on Deep Transfer Learning," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 6, pp. 3126-3134, 1 Nov.-Dec. 2022, doi: 10.1109/TCBB.2021.3128172.

[15]. K. Guo et al., "Traffic Data-Empowered XGBoost-LSTM Framework for Infectious Disease Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 1, pp. 774-785, Jan. 2024, doi: 10.1109/TITS.2022.3172206.

[16]. P. K. Sahoo, S. K. Mohapatra and S. -L. Wu, "Analyzing Healthcare Big Data With Prediction for Future Health Condition," in IEEE Access, vol. 4, pp. 9786-9799, 2016, doi: 10.1109/ACCESS.2016.2647619.

[17]. R. Wang, C. Ji, Z. Jiang, Y. Wu, L. Yin and Y. Li, "A Short-Term Prediction Model at the Early Stage of the COVID-19 Pandemic Based on Multisource Urban Data," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 938-945, Aug. 2021, doi: 10.1109/TCSS.2021.3060952.

[18]. M. Adhikari, A. Hazra and S. Nandy, "Deep Transfer Learning for Communicable Disease Detection and Recommendation in Edge Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 4, pp. 2468-2479, 1 July-Aug. 2023, doi: 10.1109/TCBB.2022.3180393.