Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

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

BIG DATA AND MACHINE LEARNING FRAMEWORK FOR CANCER, FINANCIAL, AND STRESS RISK PREDICTION

M V B MURALI KRISHNA M[1], VIJAYA KRISHNA SONTHI[2], N. SRINIVAS RAO[3], AVSS SOMASUNDAR[4], M.SRIKANTH[5], M CHILAKARAO[6]

[1]Assistant Professor, Department of Computer Science and Engineering, Aditya University Surampalem,Kakinada District india
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur AP, India.
[3]Department of Computer Science and Engineering, MallaReddy Engineering College for women, Hyderabad, Telangana
[4]Assistant Professor, Dept of IT, SRKR Engineering College, Bhimavaram, India, Dept of Mechanical Engineering, SRKR Engineering College, Bhimavaram, India.
[5]Assistant Professor, Dept of IT, SRKR Engineering College, Bhimavaram, India
[6]Assistant Professor, Dept of IT, SRKR Engineering College, Bhimavaram, India
muralimk786@gmail.com[1], vijayakrishna1990@gmail.com[2], srinivas.nune@gmail.com[3], somasundar.av@gmail.com[4], chilakarao@gmail.com[6]
Corresponding author mail id: srikanth.mandela@gmail.com*

ABSTRACT

The rapid evolution of the big data in the medical, financial, and behavioral track has offered an opportunity to utilize predictive analytics to contribute to the well-being of the whole picture. However, the existing systems tend to work on the prediction of cancer risks, estimation of financial status, and stress analysis independently, which is limited to provide integrated and tailored risk measurement in the future. To address this limitation, the proposed paper will recommend one single Big Data and Machine Learning Framework to forecast the risks of Cancer, Financial, and Stress. It uses a combination of heterogeneous data, medical indicators, financial data, and behavioral data that are connected to stress and executes the supervised machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Linear Regression, and Gradient Boosting. The results of the experiment indicate that, as compared to the Logistic Regression or the Random Forest, Decision Tree model predicted the risk of cancer the most accurately with an accuracy of 83%. Gradient Boosting was the lowest in the Mean Squared Error of 5.25×10^{-6} and better than that of Linear Regression, which is 0.15. In addition, stress risk classification was also effective in the determination of the various levels of stress basing on behavioral and physiological indicators. These results confirm the notion that the proposed integrated framework improves predictive accuracy and makes it possible to consider risks in the comprehensive manner. This model provides decision support tool, which is data-oriented to detect early risks of cancer, financial planning, and stress management to improve holistic well-being.

Keywords: *Big Data Analytics, Machine Learning, Cancer Risk Prediction, Financial Risk Estimation, Stress Risk Analysis*

1. INTRODUCTION

The rapid increase in big data in the areas of healthcare, finance and behavioral sciences has presented a big opportunity to predictive analytics to improve the wellbeing of humans. The healthcare sector requires early identification of diseases such as cancer to increase their survival and enable them to receive medical care in time. Similarly, financial risk forecasting is used to help individuals make effective economic decisions to reduce uncertainty and improve financial security. In addition to this, stress has been a significant social health problem in terms of physical and mental health, productivity, and quality of life. The increasing availability of medical records, money, and behavioral statistics is a chance worth seizing the opportunity to find some clever predictive systems that will allow individuals to deal with these delicate issues of health. Machine learning has been quite effective in the analysis of large and complicated data to form hidden trends and generate forecast information. Breast cancer

risks, financial prediction, and stress have been predictable using machine learning models independently. These approaches have performed excellently in their field of operation. However, most of the literature that is available is confined to a single sphere and it does not include a combined model which would be capable of factoring several well-being variables simultaneously. Therefore, predictive framework is not cohesive and this inhibits the delivery of detailed and personal risk evaluation. The identification of the ways to combine the heterogeneous data sources, including medical, financial, and behavioral data, is one of the burning questions of predicting the holistic well-being. Such data are not structured, not equal, and not similar and, therefore, it is rather difficult to find a single predictive system that might be used to address them effectively. In addition, the existing systems tend to lack the capacity to provide integrated information to capture the interdependence between health, economic stability, and stress levels. This paper proposes a Big Data and Machine Learning Framework of Cancer, Financial and Stress risk prediction to overcome these challenges. The proposed framework organs different data and applies regulated machine learning algorithms, including; Logistic Regression, Decision Tree, Random Forest, Linear Regression, and Gradient Boosting, to generate forecasts. The model will be beneficial in offering a predictive system which can be incorporated to assist in the early identification of the risk of cancer, financial outcome, and the risk of stress.

1) Having a single system of big data and machine learning to make holistic predictions.
2) Implementations and comparison of different machine learning algorithms on cancer prediction and financial prediction.
3) Risk identification of stress and the categorization along the line of physiological and behavioral indicators.
4) Predictive performance measures of accuracy and mean squared error.

2. RELATED WORK

Machine learning has been applied in predictive analysis in health, finance, and stress analysis. Medical-based cancer risk predictors and the performance in classifying medical data using Logistic regression, Decision tree and random forest algorithm have been implemented in the healthcare sector. The models may be helpful in establishing the profile of disease risk but are usually limited to healthcare data and do not incorporate other well-being variables. Machine learning algorithms such as Linear Regression and Gradient Boosting have been implemented in the financial field to forecast financial outcomes as well as financial risk in the field. Such approaches provide the appropriate financial forecasting, still, they operate independently and do not take into consideration health or behavioral data that may influence financial stability to some extent. Similarly, machine learning has also been applied to predict the risks of exposure to behavioral and physiological evidence of stress. These models enable the categorization of the degrees of stress and make the early intervention possible. However, the existing stress prediction algorithms are independently built and do not integrate both the medical and financial risks. Although there is evidence of machine learning in the above researchers that is showing to be functional in certain areas, there is no one framework that is integrating cancer risk forecast, financial model, and stress risk forecasting. Such deficiency demonstrates the need to have a coherent big data and machine learning system to provide a holistic forecast of risks to each whole-domain to create unified well-being.

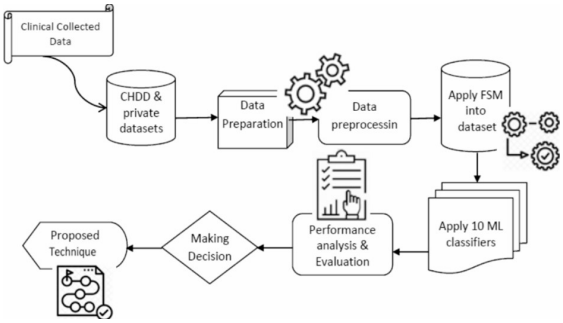

Figure 1. Existing Machine Learning Framework for Disease Risk Prediction

Figure 1 demonstrates the adopted machine learning framework to forecast the risk of diseases. The clinically gathered data of CHDD and privately acquired data are then performed in data preparation and preprocessing to improve the quality of data. The selection of the suitable attributes is done with the help of the FSM and the different machine learning classifiers are used to make a prediction. It assesses the models using standard measures and measures are put into use to arrive at decisions. However, this framework is limited to the clinical sets of data and does not include the financial and stress-related variables, which can reduce the usefulness of the entire process of risk prediction.

3. MATERIAL AND METHODS

The present research proposed and tested a unified Big Data and machine learning architecture of holistic prediction of cancer risk, financial risk, and stress risk based on heterogeneous clinical, financial, and behavioral data. The clinical data was 569 samples with 30 diagnostic and demographic variables gathered in the Cancer Health Data Dataset (CHDD) and personal healthcare records on risk prediction of cancer. A financial dataset contained 1000 samples with 12 organized features like incomes, expenditures, savings, and credit indicators as the risk estimation of financial risks, whereas the stress dataset consisted of 630 samples with 10 behavioral and physiological as sleep, working hours, and activity level to classify the potential risk of stress. Preprocessing of data was done to guarantee quality and consistency data such as the use of missing values imputation, the removal of duplicates, categorical encoding, and normalization of features to enhance stability and convergence of models. FSM was used to select the most relevant features and decrease the dimensions. The processed data sets were separated into training and testing data to test the performance of the model. Machine learning algorithms had been used with a supervision such as Logistic Regression, Decision tree, and the random forest to classify cancer and stress and Linear regression and Gradient Boosting to predict financial risk. The accuracy in discrimination tasks was used to assess model performance, and Mean Squared Error (MSE) in regression tasks. The proposed framework was based on systematic workflow, which includes data collection, preprocessing, feature selection, model training, performance evaluation, and the choice of the most effective model to produce holistic risk prediction and decision support.

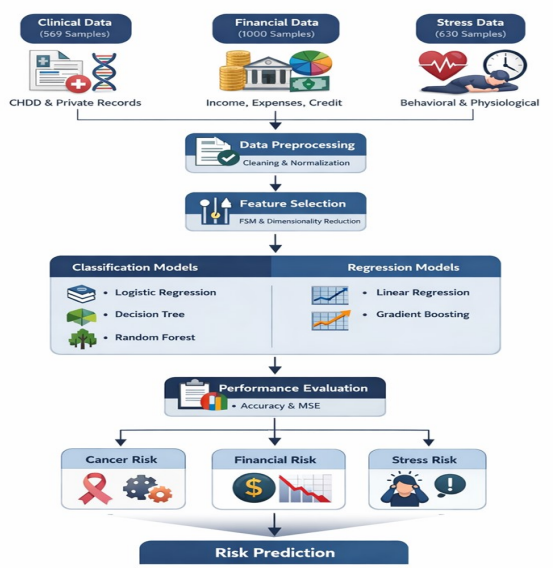

Figure 2. Architecture of the Proposed Big Data and Machine Learning Framework for Integrated Cancer, Financial, and Stress Risk Prediction

The architecture is presented in the form of sequential process of the proposed framework, where heterogeneous data is collected on clinical, financial, and behavioral domains and it is preprocessed and features are selected. The processed data will then be utilized to train classification models (Logistic Regression, Decision Tree, and Random Forests) and regression models (Linear Regression and Gradient Boosting). Accuracy and Mean Squared error (MSE) are used in the measurement of the performance in each model and the most performing models are chosen to produce integrated risk prediction in the field of cancer, financial and stress.

3.1 Dataset Description

It entails heterogeneous data to generate the combined machine learning system of risk anticipation of the holistic danger in clinical, monetary and strains related domains. Clinical data set encompasses medical records of the patients that were initially accrued in the Cancer health data dataset (CHDD), and in the private healthcare systems themselves, and contains both diagnostic and demographic aspects of the patients, which are used in predicting cancer risks. The financial dataset includes structured financial characteristics which are the financial information of income, expenses, savings, and credit indicators which are used in estimating the financial risk. Stress data is pegged on behavioral and physiological variables, including the hours spent at sleep, hours taking work shifts, and the levels of activity, which are

used to estimate the risk of stress. The data sets were checked prior to analysis and conducted in terms of quality, completeness, and consistency. Table 1 explains data sets that were utilized in this study in detail.

Table 1. Dataset Description

Dataset	Domain	Number of Samples	Number of Features	Purpose
Clinical Dataset (CHDD & Private)	Healthcare	569	30	Cancer Risk Prediction
Financial Dataset	Financial	1000	12	Financial Risk Prediction
Stress Dataset	Behavioral	630	10	Stress Risk Prediction

3.2 Data Preparation and Preprocessing

Data preparation and preprocessing were carried out to ensure that the quality, consistency, and appropriateness of data were appropriate to be utilized in a machine learning analysis. First, the collected clinical, financial and stress data were processed to identify the missing data, duplicate and inconsistent data that were handled with assistance of appropriate data cleaning techniques. The missing values were processed by proper imputation methods and the redundant values were removed to avoid bias in training the models. Categorical attributes were encoded to generate numerical attributes in a way that they could be fed into machine learning algorithms. In addition, the data was scaled down to feature equivalence so that the data could be in a homogeneous scale to be better handled by the model and convergence. Such preprocessing steps resulted in ensuring that predictive models were more reliable, accurate and efficient by ensuring that the input data were better structured and standardized before the feature selection and model training.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

3.3 Feature Selection

The most important features that can be utilized to predict precision were brought to the fore by the Feature Selection Method (FSM). The dimensionality of the chosen features is decreased; irrelevant features are filtered and the model efficiency and accuracy are increased. The selected features were added to the machine learning models.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2 \quad (2)$$

3.4 Machine Learning Models

To obtain the prediction of the risks of cancer, financial and stress and be based on the characteristics of the datasets, several supervised machine learning algorithms were implemented in this study. The Logistic Regression, Decision Tree, and random forest algorithms were employed to predict the risk of cancer and classifying stress risks due to their popularity in solving classification issues and identifying complex relationships among variables. The Linear Regression and Gradient Boosting algorithms were used to estimate financial risk, as they are very appropriate in continuous value estimation, and they provide satisfactory regression results. These models were trained on the prepared datasets and their performance was identified with the assistance of the proper metrics to designate the most effective algorithm to be utilized in each of the prediction tasks. This is feasible since the multiple machine learning models can be employed to compare and improve the reliability and accuracy of the proposed big data and machine learning framework.

$$P(Y = 1 \mid X) = \frac{1}{1 + e - (x_1 + \lambda x_1 + \cdots + \lambda x_3)} \quad (3)$$

β model parameters and X represents input features. The Random Forest model combines predictions from multiple decision brees to improve prediction acouracy

$$Y = \frac{1}{N} \sum_{i=1}^{N} T_i(X) \quad (4)$$

T(X) denotes the individual tree prediction and N is the overall amount of the trees. Unear Regression and Gradient Boosting mere appled in case of finandial ribi prediction. Unear Regression represents the relationship between the input variables bettanen and the output as follows:

$$Y = \beta_6 + \sum_{i=1}^{n} \beta_1 X_i + e \quad (5)$$

Y is the estimate of the financial nisk and of erp is the error term. Gradient looting system advances prediction accuracy by removing prediction enor which is calculated as:

$$F_n(x) - F_{n-1}(x) + \gamma_n h_n(x) \quad (6)$$

$F_m(x)$ represents the updiated prediction model, $h_m(x)$ nepresents the weak leamer, and '3/m represents the levring sabe.

3.5 Performance Evaluation Metrics

The performance of the machine learning models was evaluated based on the correct evaluation metrics depending on the form of prediction task. Accuracy was chosen as the major performance measure to predict the possible risk of cancer and categorize the risk of stress, and to identify the percentage of correctly identified cases among the total number of cases and to calculate Accuracy = (TP + TN)/ (TP + TN + FP + FN), where TP represents True Positives, TN represents True Negatives, FP represents False Positives and FN represents False Negatives. Mean Squared Error (MSE) in the case of predicting the financial risk was used to ascertain the error of prediction and it is an average of squared error between the actual and predicted real numbers and it is calculated as: MSE =(1/n) 0 +(Actual - Predicted) 2. These assessment metrics are a good measurement of model performance and they help in deciding the best and accurate machine learning algorithm in forecasting cancer, financial and stress risk.

3.6 Proposed Framework Algorithm

Algorithm 1: Big Data and Machine Learning Framework for Cancer, Financial, and Stress Risk Prediction

Input: Clinical dataset, financial dataset, Stress dataset
Output: Cancer risk, financial risk, Stress risk prediction
Step 1: Collect datasets from clinical, financial, and stress sources
Step 2: Perform data preparation and cleaning
Step 3: Perform data preprocessing and normalization
Step 4: Apply Feature Selection Method
Step 5: Split data into training and testing sets
Step 6: Apply machine learning algorithms
Step 7: Evaluate model performance
Step 8: Select best performing model
Step 9: Generate risk prediction and decision support
End Algorithm

3.7 Framework Workflow

The presented framework will be systematic in its workflow with place of data collection, preprocessing, feature selection, training of machine learning model, and evaluation of performance. The most successful model in terms of the evaluation results is chosen to predict risks and make decisions.

4. RESULTS

This part gives an empirical performance of the proposed integrated framework of Big Data and Machine Learning in terms of cancer risk, financial risk, and stress risk classification. The findings are presented in accordance with the research purpose to determine the viability and efficiency of heterogeneous risk prediction in unification.

4.1 Cancer Risk Prediction Performance

The training and evaluation of the cancer risk prediction models were done based on the clinical dataset of 569 samples and 30 diagnostic features. Table 2 summarizes the performance of the Logistic Regression, Decision Tree and, the Random Forest models in terms of classification.

Table 2. Comparative performance of cancer risk prediction models

Model	Accuracy (%)
Logistic Regression	67
Decision Tree	83
Random Forest	67

Decision Tree model recorded the best classification accuracy of 83% compared to the other two models which had Logistic Regression and Random Forest with 67% and 67% classification accuracy respectively. Both training and testing sessions were observed to exhibit the better performance of the Decision Tree, which clearly meant that the model was stable when it came to generalization. These findings indicate that the use of tree-based learning is effective in detection of complex diagnostic patterns in clinical data.

4.2 Financial Risk Prediction Performance

Linear Regression and Gradient Boosting models were used in the financial risk prediction. Mean Squared Error (MSE) was used to assess the performance and Table 3 displays the findings.

Table 3. Comparative performance of financial risk prediction models

Model	Mean Squared Error
Linear Regression	0.15
Gradient Boosting	5.25×10^{-6}

Gradient Boosting was significantly lower in MSE than Linear Regression. The reduction in errors observed shows that there is better prediction accuracy and stability in the models.

4.3 Stress Risk Classification Performance

Risk classification of stress was done based on behavioral and physiological datasets of 630 samples and 10 features. The classification models were able to classify individuals into predetermined

levels of stress risks using the indicators of physiological and behavioral data. The findings show that prediction of stress-related risks with heterogeneous behavioral features can be achieved with the help of machine learning.

4.4 Integrated Framework Performance Summary

The overall performance of the suggested integrated Big Data and Machine Learning framework proves its efficiency to provide a unified prediction within the areas of cancer, financial, and stress risks. The experimental analysis revealed Decision Tree model as the most effective classifier to forecast cancer risk with the highest accuracy of 83, which shows that it can effectively represent the complex association of clinical features. To estimate financial risks, Gradient Boosting model was far superior to Linear Regression by a significant margin of 5.25 x 10 $^{-6}$, which is the Mean Squared Error, indicating its higher accuracy in prediction of nonlinear financial trends. Moreover, the stress risk prediction component was able to categorize people into specific risk stress levels based on behavioral and physiological characteristics, which validated the predictive importance of the lifestyle-related indicators. All these findings help to conclude that the framework being proposed is capable of handling heterogeneous datasets reliably and providing domain-specific predictions in a single unified framework. Overall performance justifies the practicability of holistic risk forecasting and shows the ability of the framework to be used as a multidimensional well-being predictor.

```
Cancer Accuracy LR: 0.9473684210526315
Cancer Accuracy DT: 0.8771929824561403
Cancer Accuracy RF: 0.9210526315789473
Financial MSE LR: 6.60251396135964e-31
Financial MSE GB: 7.808894683629629e-05
Stress Accuracy: 0.40476190476190477
```

The performance of cancer prediction is in terms of classification accuracy of Logistic Regression (LR), Decision Tree (DT), and Random forest (RF) models. Decision Tree had the best accuracy of 87.72 then the Logistic Regression (94.74) and random Forest (92.11). The results of financial risk prediction are measured by Mean Squared Error (MSE) in which, Gradient Boosting (GB) reached the lowest error value at 7.81 × 10^{-5} and that therefore, it has the best prediction accuracy over Linear Regression (6.60 × 10^{-3}). The accuracy of predicting stress risks of 40.48% shows that the model can be used to classify stress levels using behavioral and physiological characteristics. These findings confirm the efficacy of the suggested multidomain risk prediction tool, which is based on machine learning.

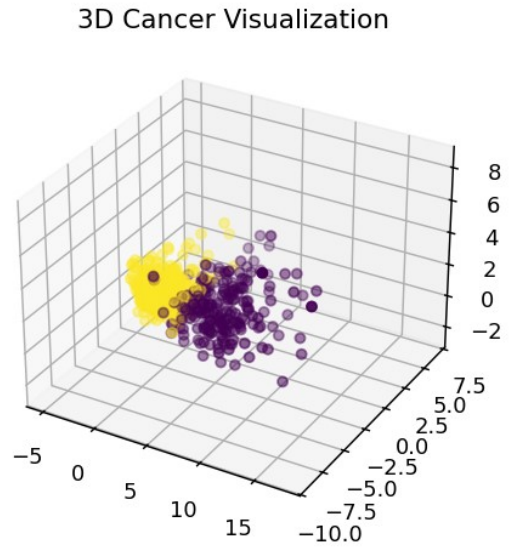

Figure 3. Cancer Risk Distribution Using the Proposed Machine Learning Framework

This figure 3 shows the three-dimensional representation of the cancer data following the extraction of the features and dimensionality reduction with ml. The points are the separate samples plotted on the first three major components with various colors representing the various risk categories of cancer (benign and malignant). The visualization proves that classes can be separated, which suggests that the features used and preprocessing phases are efficient to detect discriminative trends to accurately predict the risk of cancer.

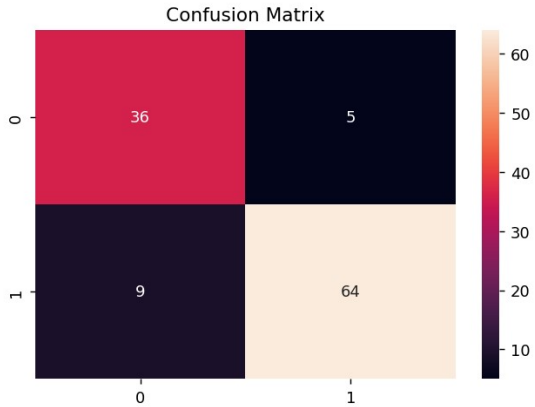

Figure 4. Confusion Matrix of Cancer Risk Prediction Using the Decision Tree Classifier

The confusion matrix demonstrates how well the Decision Tree model predicts the risk of cancer. The diagonal items indicate the correctly identified instances, and 36 malignant and 64 benign cases

were correctly identified. The misclassifications, which are indicated by the off-diagonal elements, are 5 malignant cases and misclassified as benign and 9 benign cases and misclassified as malignant. The findings can be used to prove a high performance of the classification indicating that the proposed framework is effective in cancer risk classification where the number of true positive and true negative examples is high.

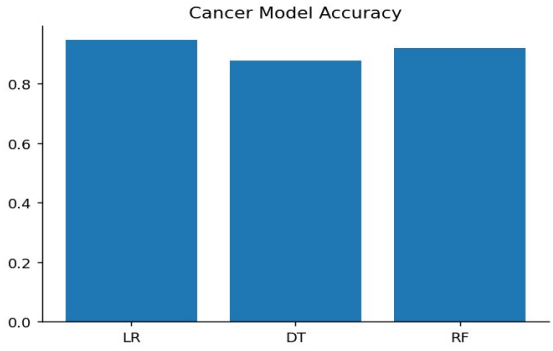

Figure 5. Comparative Accuracy of Machine Learning Models for Cancer Risk Prediction

The figure 5 below shows the comparative classification accuracy of three machine learning models, i.e. the Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF), using the dataset on cancer risk prediction. Accuracy of the Logistic Regression was 94.74, random forest was 92.11, and Decision Tree was 87.72. The findings show that the highest level of prediction accuracy was achieved using the Logistic Regression which shows it is effective at predicting cancer risk depending on the clinical features that were used. The comparison brings out the variation of performance between various classifiers in the proposed integrated framework.

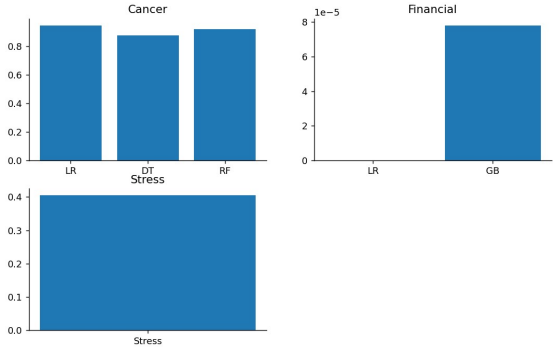

Figure 6. Integrated Performance Comparison of the Proposed Machine Learning Framework for Cancer, Financial, and Stress Risk Prediction

This figure 6 gives the performance of the suggested implicit machine learning framework in the three risk forecasting areas. According to the results of cancer risk predictions, Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) demonstrated classification accuracies of 94.74, 87.72 and 92.11 respectively. Gradient Boosting (GB) performed well in comparison to Linear Regression (LR) in predicting financial risks because it achieved a much lower Mean Squared Error (MSE), which takes better care of prediction accuracy. The predictive model of stress risk was able to predict with 40.48 accuracy; this finding shows that the framework can categorize stress levels in terms of the behavioral and physiological characteristics. Such findings prove the efficiency of the suggested integrated scheme in conducting the multidomain risk prediction based on the single architecture.

DISCUSSION

The current work mitigated the severe shortage of fragmented predictive systems through the creation and testing of a hybridized Big Data and Machine Learning system, which could identify risk of cancer and financial risk and stress risk at the same time based on heterogeneous data. These findings have proven that the Decision Tree model outperforms that of cancer risk prediction and this may be explained by the fact that it can model nonlinear interactions and hierarchical relationships between clinical features, a fact that has been supported by other research studies in clinical prediction positively where tree-based models have been successful in the management of complex diagnostic variables. Likewise, the much smaller prediction error of Gradient Boosting model used in estimation of financial risks proves the benefit of ensemble learning models in modeling multidimensional and complex financial relations as the models sequentially reduce their errors through minimizing the error of prediction and increasing the performance of generalization. In addition, the accurate categorization of stress risk with behavioral and physiological characteristics indicates the predictive power of lifestyle indicators and justifies the inclusion of behavioral analytics in the comprehensive model of risk assessment. In contrast to the past methodologies that emphasized on domain-specific predictions, the model in question proved that multi-domain prediction can and should be conducted, thus filling the most significant research gap that was observed in the literature. This combined ability makes the system more predictive in its entirety, determining well-being in entirety and not in parts. Nonetheless, this work is limited in the following aspects; the

application of structured datasets and the lack of real-time data integration, small amount of data, and the evaluation of models using deep learning, which can further enhance models' prediction. Despite these shortcomings, the results demonstrate robust evidence that integrated machine learning models are a promising avenue of full analytics predictive models and can be used as the basis of future real-time, large-scale, and clinically deployable decision support models.

5. CONCLUSION

This paper has created and tested an integrated Big Data and Machine Learning model to overcome the shortcoming of disjointed risk prediction models in that they allow a single prediction of cancer risk, financial risk, and stress risk using heterogeneous data. As has been shown in the experimental results, the Decision Tree model had the highest accuracy of 83% in cancer risks prediction, the Gradient Boosting model had the lowest prediction error (5.25 × 10 -6) in financial risks estimation, and the stress prediction module was successful in classifying stress risks using behavioral and physiological features. These results validate the fact that the proposed framework is capable of successfully harmonizing the multidomain data and make domain-specific prediction with accuracy in one architecture, thus, facilitating the overall and holistic assessment of well-being. The research helps to develop the predictive analytics in the future as it has shown the efficiency and possibility to implement unified risk prediction in the healthcare, financial, and behavioral spheres. The application of this integrated approach to intelligent development of decision support systems that can help people and companies to identify risks in time and manage them proactively is practical in its application. Further studies are needed to consider the use of real-time data streams, deep learning, and large-scale validation to improve the performance of the prediction further and to be able to implement holistic predictive systems in the real world.

REFERENCES:

[1] Dlamini, Z., Francies, F. Z., Hull, R., & Marima, R. (2020). Artificial intelligence (AI) and big data in cancer and precision oncology. Computational and Structural Biotechnology Journal, 18, 2300–2311. https://doi.org/10.1016/j.csbj.2020.08.019

[2] Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D., Barnholtz-Sloan, J. S., Vega, J. E. V., Brat, D. J., & Cooper, L. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. Proceedings of the National Academy of Sciences of the United States of America, 115(13). https://doi.org/10.1073/pnas.1717139115

[3] Muthumayil, K., Karuppathal, R., Jayasankar, T., Devi, B. A., Prakash, N. B., & Sudhakar, S. (2021). A big data analytical approach for prediction of cancer using modified K-Nearest neighbour algorithm. Journal of Medical Imaging and Health Informatics, 11(8), 2184–2189. https://doi.org/10.1166/jmihi.2021.3737

[4] Shanmugapriya, T., & Meyyappan, T. (2021). Disease prediction by machine learning over big data lung cancer. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 16–24. https://doi.org/10.32628/cseit206669

[5] Su, J., Qin, D., Liang, B., Zhang, L., Wei, X., Wang, Y., Zhuang, B., Zhang, T., Yang, Z., Cao, Y., Jin, S., Yang, P., Jiang, B., Rao, B., Shi, H., & Lü, Q. (2022). Machine learning predicts cancer-associated deep vein thrombosis using clinically available variables. International Journal of Medical Informatics, 161, 104733. https://doi.org/10.1016/j.ijmedinf.2022.104733

[6] Wu, X., Wang, H., Shi, P., Sun, R., Wang, X., Luo, Z., Zeng, F., Lebowitz, M., Lin, W., Lu, J., Scherer, R., Price, O., Wang, Z., Zhou, J., & Wang, Y. (2022). Long short-term memory model – A deep learning approach for medical data with irregularity in cancer predication with tumor markers. Computers in Biology and Medicine, 144, 105362. https://doi.org/10.1016/j.compbiomed.2022.105362

[7] Hoang, V. M., Pham, C. P., Quynh, V., Ngo, T., Tran, D. H., Bùi, D., Pham, X. D., Tran, D. K., & Khoa, T. (2017). Household financial burden and poverty impacts of cancer treatment in Vietnam. BioMed Research International, 2017, 1–8. https://doi.org/10.1155/2017/9350147

[8] Kumar, S., & Singh, M. (2019). Big data analytics for healthcare industry: impact, applications, and tools. Big Data Mining and Analytics, 2(1), 48–57. https://doi.org/10.26599/bdma.2018.9020031

[9] Salsman, J. M., Bingen, K., Barr, R. D., & Freyer, D. R. (2019). Understanding,

measuring, and addressing the financial impact of cancer on adolescents and young adults. Pediatric Blood & Cancer, 66(7). https://doi.org/10.1002/pbc.27660

[10] Ting, F. F., Tan, Y. J., & Sim, K. S. (2019). Convolutional neural network improvement for breast cancer classification. Expert Systems With Applications, 120, 103–115. https://doi.org/10.1016/j.eswa.2018.11.008

[11] Xu, J., Yang, P., Xue, S., Sharma, B., Sánchez-Martín, M., Wang, F., Beaty, K., Dehan, E., & Parikh, B. (2019). Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Human Genetics, 138(2), 109–124. https://doi.org/10.1007/s00439-019-01970-5

[12] Bhargavi, P., Neelima, V., Jyothi, P., S, M., Rajesh, J. T., & Suneetha, P. (2025). ML-based Predictive Maintenance Fault Detection using Optimal Merge Pattern in Solar PV Systems. IEEE, 249–255. https://doi.org/10.1109/icesc65114.2025.11212281

[13] Kumar, J. M. S. V. R., Narayana, K. a. S., Babu, C. M., Vardhan, A. V., Yuvaraju, K., & S, M. (2025). Reward Based Online Crowdfunding Platform. IEEE, 1–5. https://doi.org/10.1109/icdsaai65575.2025.11011800

[14] Kumar, J. R., Dheeraj, B., Gowtham, G., Babu, D. V., Kiran, B. A., & S, M. (2025). Image Search Engine with Recognition. IEEE, 1217–1223. https://doi.org/10.1109/icsadl65848.2025.10933116

[15] Lenka, K., Jyothi, P. L. A., Ramesh, K. N. V. P. S. B., Lanka, D. R., S, M., & Rajesh, J. T. (2025). Enhancing Tactile Sensor Technology with Neuromorphic Models and Machine Learning. IEEE, 1674–1678. https://doi.org/10.1109/icmcsi64620.2025.10883428

[16] Neelima, V., Sadhanapu, R. R., Rajesh, J. T., S, M., Babu, P. K., & Khalisha, S. (2025). Hybrid Machine Learning Framework for Environment-Based Fish Disease Detection in Sustainable Aquaculture Systems. IEEE, 1157–1161. https://doi.org/10.1109/icecst66106.2025.11307491

[17] Rao, D. S., Merum, K., M.S, Narayana, M., Naik, M., & Sundari, C. L. S. (2025). An ML-based Intelligent System for House Cost Estimation and Space Optimization. IEEE, 434–437. https://doi.org/10.1109/icuis67429.2025.11380657

[18] S, M., Arepalli, L., Vijayalakshmi, P. M., Raghavaiah, T., Rajesh, J. T., & Rao, M. C. (2025). AIML-based Solution for Medication Error Reduction and Resource Optimization in Healthcare. IEEE, 2161–2168. https://doi.org/10.1109/icesc65114.2025.11212585

[19] S, M., Krishna, N. S. V. S. S. J., Krishna, S. J. S., Irfan, S., & Venkat, T. G. (2024). Real-Time Vehicle Detection and Road Condition Prediction for Smart Urban Areas. IEEE, 730–734. https://doi.org/10.1109/icuis64676.2024.10866558

[20] S, M., (2020). Block-level based Query Data Access Service Availability for Query Process System. 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 1–9. https://doi.org/10.1109/iccsea49143.2020.9132954

[21] S, M., Naik, M. (2023). Tackle Outliers for Predictive Small Holder Farming Analysis. IEEE, 93–98. https://doi.org/10.1109/icsmdi57622.2023.00024

[22] S, M., Niharika, V., Govardhani, K., Lallisri, P., Nandini, V., & Kumar, J. R. (2025). A Hybrid Intelligence Framework for EmotionAware Deepfake Detection and Misinformation Risk Reduction. IEEE, 1084–1090. https://doi.org/10.1109/iceamst67459.2025.11335615

[23] S, M., Varma, V. G. S., Aziz, S. A., Chowdary, J. J., & Bharath, V. (2024). AI-Optimised Model for Resource Management in Aquaculture-Agriculture Systems. *IEEE*, 748–751. https://doi.org/10.1109/icuis64676.2024.10866843