Innovative Diagnostic Strategies for Identifying High-Risk Patients Using Laboratory Parameters

Abstract: Identifying high-risk patients early in an advanced treatment planning setting is crucial and is comparable to addressing diseases such as cancer, which pose significant threats to human lives. New diagnostic techniques can improve health outcomes and reduce treatment costs, and lead to more efficient resources utilize in the future. The research developed a risk classification framework for high-risk patients using deep learning methodologies based on laboratory parameters. These parameters include outputs directly obtained from patient record data, lab test results, demographic information, and clinical outcomes. Independent Component Analysis (ICA) was employed to preprocess the data, which had already undergone data cleaning and normalization. The hybrid approach, incorporating Wolf Pack Search with flexible parameters and Deep Belief Network (FlexiWPS-DBN) combines the best features of both models. The approach presents a more accurate and comprehensible method for identifying high-risk patients based on laboratory parameters. The findings highlight composite blood tests, documented, provide improved accuracy compared to traditional algorithms. The proposed method achieved 92.2% accuracy, 91.5% precision, 91.6% recall, and a 90.3% F1 score. The specialized, parameter-driven, lab tool significantly increases diagnosis accuracy and reducing the strain on the healthcare delivery system. By enabling proactive detection of at risk individuals, it could serve as key component of cancer management program for populations.

Keywords: Diagnostic Strategies; High-Risk Patients; Laboratory Parameters; Cancer; Flexible Wolf Pack Search driven Deep Belief Network (FlexiWPS-DBN)

Introduction

Diagnostic strategies are tools and systems that help with early discovery of the risk factors for disease, making possible early intervention when necessary. Recently, testing to assess clinical laboratory findings and patient data has made use of machine learning (ML) and deep learning (DL) [1 & 2]. Very traditional diagnostic tests, like blood counts and metabolic panels, are widely used in clinics. Besides this, innovation has made new advances in diagnostic analytics such as developing feature extraction and optimization algorithms. This is where personalized healthcare, as a high-effectiveness program, comes about; it identifies treating individuals who are at large risk, thus increasing treatment, lowering costs of healthcare, and optimizing internal resources. The diagnosis strategies for high-risk patients can include taking comprehensive medical histories, risk factor assessments, and physical examinations [3], Imaging, as well as laboratory testing, genetic screening; provide further scope in order to confirm all that is the point of differentiation. The risk profiling in patients did not just end with profiling but stood extended into the relating resource utilization formulation and allocation modeling. Additionally, the application of predictive analytics and artificial intelligence (AI) helps in efficient high-risk profiling. Wearables and collaborative models with improved outcomes in the realm of digital health risk management are being encouraged to develop more individually tailored routes for interventions [4 & 5]. A prediction from laboratory markers recognizing high-risk patients with routine clinical blood tests, biomarkers, or specific biomarkers can help in estimating various kinds of bad outcomes or progression of disease in patients. The technique enhances the information derived from clinical sources by incorporating patient history,

thereby increasing reliability of accurate risk stratification [6]. AI and ML are increasingly used to analyze large datasets, revealing patterns enabling the fast identification of high-risk patients and timely interventions. However, continuous monitoring and regular updating of diagnostic models are crucial to ensure that they remain relevant as patients' progress or change [7 & 8]. The prevention issues in laboratory test-based diagnosis strategies, such as unsatisfactory completion, poor data quality, variations in the test results, and limited applicability across population checking from different countries requires comprehensive measures. Potential solution includes robust data preprocessing, standardized protocols for laboratory tests and training models with high-quality diverse datasets to ensure broad generalization and reliability. The research aims to develop diagnostic strategies for identifying high-risk patients using a DL-based risk classification framework applied to laboratory parameters, focusing on individuals at elevated risk for various types of cancer.

Literature Reviews

The investigation found ML can improve primary care risk-assessment tools by identifying people at high risk for esophageal cancer through electronic medical information [9]. Information from the United States General Practice Research Database was utilized to create five probabilistic ML classifiers. The ML models identified 11% more cancer patients than the existing methods. The conclusion ML can improve primary care cancer risk assessment tools and possibly raise survival rates and facilitating more precise identification and rapid therapy. To identify risk factors and biomarkers associated with Immune-related Adverse Events (irAEs) in patients on Immune Checkpoint Inhibitors (ICI), a comprehensive review of the literature was conducted in [10]. Biomarkers, tumor-specific and agent-specific risk factors, medical history, social background, and demography were included among the risk factors. The most important factors included medication history, pre-existing conditions, social history, and demography. Thyroid-stimulating hormone, serum albumin, cytokines, and the neutrophil-lymphocyte ratio were capable biomarkers; however, further research was needed to validate effectiveness. Understanding biomarker was crucial for making informed clinical decisions regarding ICI therapy.

Thorough an analysis of pancreatic cancer trends, risk variables, prediction models, screening methods, and prognosis was conducted in [11]. Personal attributes, lifestyle, and condition all represent risk. Validation was required for any prediction models designed for diabetes developing very recently or a family history of pancreatic cancer. Advances in screening remain very few, especially in high-risk groups. New preventative and therapeutic approaches were proposed.DL models were developed to distinguish between high-risk and low-risk cancers [12]. AI-biopsy offered a reliable and data-driven way to use Magnetic Resonance Imaging (MRI) to estimate the likelihood of malignancy, potentially avoiding needless biopsies. to use an MR image data set tagged with histological information to develop an AI biopsy model for the early detection of prostate cancer. The model using 228 internal and 172 external databases of histology data of approximately 400 men suspected of having prostate cancer. The investigation discussed the topic concerning the power of AI in assisting the early stages of cancer abnormalities. The research examined the technology's function in diagnosing and treating cancer [13]. Technologies driven by AI identify genetic mutations, diagnose conditions, forecast illness risk, and customize patient care. The research highlighted the potential of AI in revolutionizing medicine,

particularly oncology, by enabling real-time collaboration and accelerating the development of new medicines.

The investigation examines new uses of DL in oncology, specifically through transcriptome, methylation, and genomic data. Given its use in healthcare and the increasing amount of data available in cancer biology, research into DL potential applications in cancer biology has accelerated. By making predictions in massive datasets, artificial neural networks (ANN) subset of AI, were utilized to find patterns in DL [14]. It discussed the limitations and challenges of precision oncology and provided solutions for potential therapeutic applications. It used a multidisciplinary approach combined ML and metabolomics to identify plasma compounds from Chinese patients as lung cancer diagnostic biomarkers [15]. A specific combination of six metabolic biomarkers was identified to distinguish between patients with stage one lung cancer and healthy individuals in the investigation, which comprised 110 patients with the disease and 43 healthy individuals. For the early detection of lung tumors, Naïve Byes (NB) was proposed in order to increase the feasibility of blood-based screening and extend it to other cancers.

Key Contributions

- Various data sources, including patient files, laboratory test results, demographic data, and clinical outcomes, are combined to create a comprehensive dataset for predicting of cancer risks.
- Data pre-processing involves cleaning, normalization, and feature extraction using ICA resulting in reliable input data for the model and enhancing the quality of input data usage.
- The proposed FlexiWPS-DBN model was developed to optimize and adaptDL approaches designed to identify high-risk patients effectively.

Methodology

Figure 1 describes the research design which involves patient records, laboratory test results, demographic details, and clinical outcomes. The data is subjected to cleaning and normalization to achieve quality before preprocessing. Features are extracted relevant to the preprocessed data using ICA. A developed model called WPS-Flexi-DBN is employed for classification of cancer high-risk patients from laboratory parameters.

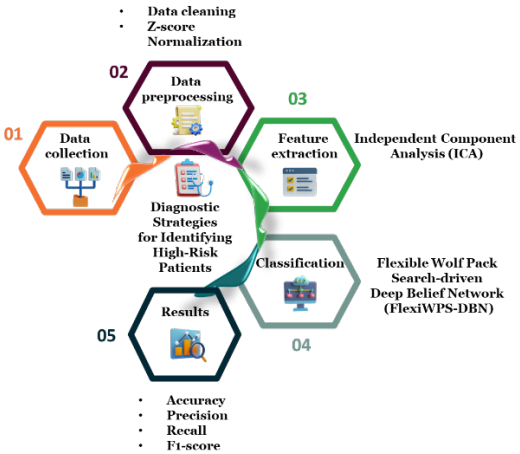

Figure 1. Research Design for Diagnostic Strategies for Identifying High-Risk Patients

Data Collection

The data was sourced from the open-access Kaggle website: https://www.kaggle.com/datasets/programmer3/cancer-risk-stratification-using-lab-parameters. The dataset "Cancer Risk Stratification Using Lab Parameters" would be very helpful in developing and evaluating the diagnostic algorithms for cancer early detection. It comprises information regarding a patient's demographics, test results, and cancer biomarkers. Key laboratory parameters across the datasets are complete blood counts, metabolic panels, and particular cancer biomarkers like CA-125, PSA, and CEA. Further, the dataset gives details of the levels of risk of the patients, the stages of cancer, results of diagnosis, treatments available, and survival status.

Data preprocessing

It involves two crucial steps in preprocessing data: it includes data cleaning and Z-score normalization. Data cleaning is meant to remove conflict and missing values present in the set of data so that it causes correct analytical procedures. The last step is the Z-score normalization, which normalizes data on a standardized scale, thus facilitating easier comparison and improving the capability of the diagnostic procedure to detect the most serious cases.

Data Cleaning

The most vital step in building diagnostic strategies that identify high-risk patients is cleaning data, where missing values arising from human mistakes, equipment malfunction, or even inconsistent or absent data are restored. The approach replaces anomalous values to ensure the production of reliable DL models for predicting at-risk patients based on laboratory parameters.

Z-Score Normalization

Z-score normalization is a statistical technique applied to the process of dealing with data outliers and standardizing features for diagnostic purposes. It makes the feature values using mean and standard deviation, thereby ensuring comparability across variables. This z-score normalization further improves the efficiency of diagnostic procedures in identifying serious cases. The transformation is performed using the following equation (1).

$$u' = \frac{u-\mu}{\sigma} \quad (1)$$

Where σ is the defined feature's standard deviation and μ is the feature's mean value. Values are equal to the mean are mapped to zero when the Z-score normalization method is applied; values are higher or lower than the mean are shown as positive or negative integers, respectively. By ensuring the attributes are standardized, the approach facilitates the identification of high-risk patients using the normalized data.

Feature Extraction

To diagnose high-risk patients based on laboratory parameters,ICA is employed to extract critical features by identifying independent components from observed data. The method highlights patterns crucial for distinguishing high-risk individuals. Given m linear mixtures measured by transducers, $w_1(s), w_2(s), \ldots\ldots, w_m(s)$ and m independent source signals $t_1(s), t_2(s), \ldots\ldots, t_m(s)$, the relationships can be expressed in equation (2).

$$w_i(s) = b_{i1}t_1 + b_{i2}t_2 + \cdots., b_{im}t_m, i = 1,2, \ldots. m \quad (2)$$

In the matrix from equation (3).

$$W = AS \tag{3}$$

Where,

$W = [w_1(s), w_2(s), \ldots\ldots, w_m(s)]^S$represents the vector of observed mixtures.

$T = [t_1(s), t_2(s), \ldots, t_m(s)]^S$represents the vector of the independent source signal.

The latent independent components $t\ (s)$ are estimated through equation (4 & 5).

$$X = B^{-1} \tag{4}$$

$$T = XW \tag{5}$$

Non-Gaussianity, as per the Central limit theorem, helps to isolate independent components by maximizing non-Gaussianity in the fast ICA algorithm. The method enables more accurate feature extraction and identification of high-risk based on laboratory data.

Diagnostic Strategies for Identifying High-Risk Patients using Flexible Wolf Pack Search-driven Deep Belief Network (FlexiWPS-DBN)

The FlexiWPS-DBN approach uses a DL model to predict patient risk based on laboratory parameters. The optimization technique through FlexiWPS fine-tunes the DBN and further enhances its high-risk detection capability. The strategy makes the diagnosis much more accurate; it is built on data-driven insights and an optimal search mechanism for performance.

Deep Belief Network (DBN)

In the context of diagnosing high-risk patients advanced models, like DBN, are widely used to analyze complex medical data. A DBN is a generative model thatworks by stacking restricted Boltzmann machines (RBM) layer by layer to extract significant features from medical records or diagnostic data. The model learns the joint probability distribution over all hidden and visible layers. The DBN can be represented by the following joint probability in equation (6).

$$o(\gamma, g^1, g^2, \ldots\ldots, g^m) = o(\gamma|g^1)o(g^1|g^2)\ldots.(g^{m-1}|g^m) \tag{6}$$

Where γ represents the visible layer, and $g^1, g^2, \ldots\ldots, and\ g^m$ are the hidden layers. The conditional distribution between the hidden layers is expressed by $lo(g^{m-1}|g^m)$. Each RBM is defined by a visible layer $d = (d1, d2, \ldots dn)$ and a hidden layer$g = (g_1, g_2, \ldots\ldots, g_k)$, where the energy function $F(\gamma, g; \theta)$ is used to evaluate the state of the system in Equation (7).

$$F(\gamma, g; \theta) = -\sum_{j=1}^{k}\sum_{i=1}^{n}\omega_{ji}g_j\gamma_i - \sum_{i=1}^{n}a_i\gamma_i - \sum_{j=1}^{k}d_jg_j \tag{7}$$

Where ω_{ji}isthe weight between the visible and hidden layers and d_j, $a_i$is the biases. To calculate the conditional probabilities of hidden and visible units in equations (8 & 9).

$$o(\gamma_i|g; \theta) = \sigma\left(\sum_{j=1}^{k}\omega_{ji}g_j + d_j\right) \tag{8}$$

$$o(g_j|\gamma;\theta) = \sigma\left(\sum_{i=1}^{n} \omega_{ji}\gamma_j + a_i\right) \quad (9)$$

Where $\sigma(w) = 1/(1+\exp(-w))$is the sigmoid function. The learning process for RBM involves updating using contrastive divergence, is represented in equations (10 - 12).

$$\omega_{ji}^{(s)} \leftarrow \omega_{ji}^{(s-1)} + \eta\left(o(g_j|\gamma_j;\theta)\gamma_j - o(g_{j+1}|\gamma_{j+1};\theta)\gamma_{j+1}\right) \quad (10)$$

$$d_j \leftarrow d_{j-1} + \eta(g_j - g_{j+1}) \quad (11)$$

$$a_i \leftarrow a_{i-1} + \eta(\gamma_j - \gamma_{j+1}) \quad (12)$$

These equations help in identifying high-risk patients by efficiently learning complex patterns from patient data.

Flexible Wolf Pack search (FlexiWPS)

Patient risk position Update: The patient's risk status is represented by the beta patient's location in the health parameter space. To enhance diagnostic accuracy, the beta patient's location updates during diagnosis phases is expressed in equation (13).

$$w_{id}^{o} = w_{id} + \sin\left(2\pi \times \frac{o}{g}\right) \times step_b^c \quad (13)$$

Where,

w_{id}^{o} represents the updated risk position of the beta patient id,

$g$is the number of diagnostic paths or directions the beta patient explores,

$step_b^c$ is the size for exploring health factors.

Where g is the number of paths a beta wolf takes while exploring. The value of$g$is often 4. The computation reveals there are only two values available for updating the Beta wolf's position.

$$\begin{aligned} w_{id}^{1} &= w_{id} + step_b^c \\ w_{id}^{3} &= w_{id} - step_b^c \\ w_{id}^{2} &= w_{id}^{4} = w_{id} \end{aligned} \quad (14)$$

When beta wolves travel in multiple directions, modification makes sure the initial position value can only be the same.

Adaptive step size for risk assessment: As the patient progresses through diagnostic phases, the step size should adapt to the proximity of risk. The updated step size is shown in equation (15).

$$w_{id}^{l+1} = w_{id}^{l} + \frac{(1-l)}{l_{max}} \times step_a^c \times \frac{(h_c^l - w_{id}^l)}{|h_c^l - w_{id}^l|} \quad (15)$$

Where,

w_{id}^{l}represents the patient's risk state at iteration l,

$h_c^l$is the current health threshold,

$l_{max}$is the total number of diagnostic iterations.

Modifying the traditional FlexiWPS for these diagnostic steps helps refine the process of identifying high-risk patients by adaptively exploring various health factors and dynamically adjusting the focus of the diagnostic steps.FlexiWPS-DBN is a diagnostic framework that uses DL to identify high-risk patients. The powerful feature learning of DBN combined with the optimization capabilities of the FlexiWPS algorithm make this hybrid model the best in analyzing large-scale high-dimensional healthcare datasets and extracting features to predict the risk levels of the patients. Inspired by the behavior of wolves during hunting, WPS fine-tunes the parameters of the network to improve the capacity of learning complex patterns within medical data. Its accuracy, adaptability, and robustness make it a potential candidate for the prediction of outcomes such as disease progression or responses to treatments. FlexiWPS-DBN is amenable to adaptation according to specific diagnostic scenarios or datasets. It provides reliable, data-driven insights in the early intervention phase and essentially adaptive care strategies for high-risk patients.

Results

The research is based on high-risk patient diagnostic strategies with the help of Python 3.10, supporting TensorFlow 2.8 and Keras. It is executed on a Windows 10 machine to improve performance by using Python 3.10. The suggested FlexiWPS-DBN method is compared with other existing methods such as SVM and XGBoost [16] for diagnosis.

Correlation matrix

Figure 2 showsacorrelation matrix linear relationship between clinical and demographic variables with values ranging from -1 to 1. Most variables exhibit weak correlations (close to 0), indicating minimal linear dependent between the values. Strong self-correlation is observed along the diagonal (Value=1) for each variable. Notable variables such as WBC, RBC, and CA_125 show no significant relationship with others.

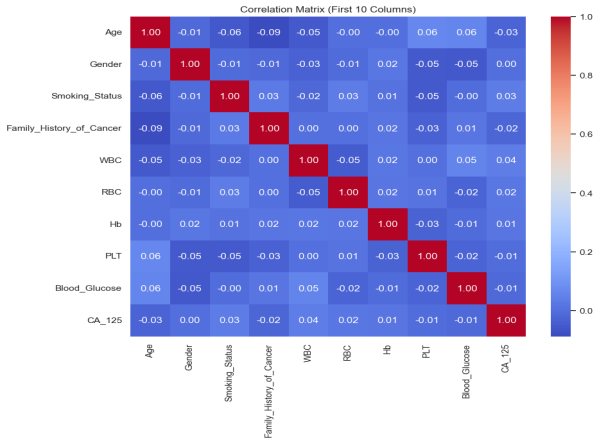

Figure 2. Correlation matrix for diagnostic strategies

Confusion Matrix

Figure 3 shows confusion matrix toevaluate the performance of a binary classification model used for diagnostic strategies. It displays four outcomes: True Negatives (161), where the model correctly classified "No" instances, and True Positives (139), where it correctly classified "Yes" instances. There are no False Negatives (0), meaning the model did not incorrectly classify any "Yes" cases as "No," and no False Positives (0), indicating no "No" cases were misclassified as "Yes." Every prediction in the matrix matches the actual results exactly; it represents a flawless classification model.

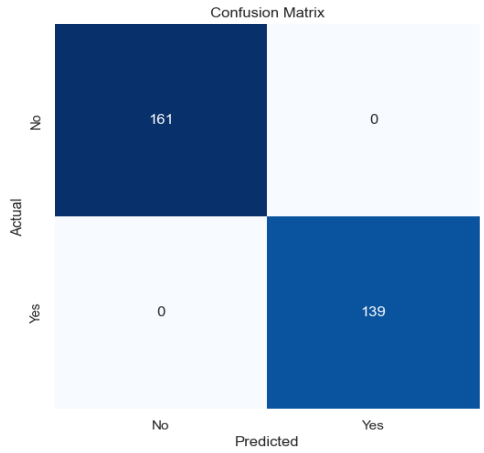

Figure 3. Confusion matrix for Binary classifier

Feature Importance

Figure 4 visualizes the importance of different features in predicting cancer risk. Key biomarkers, such as PSA, CA-125, Hb, RBC, and CEA, have the highest importance, indicating a strong contribution to the model's predictions. Other factors like age, blood glucose, and demographic variables (e.g., gender, smoking status, and family history) show comparatively lower significance. These insights help prioritize critical features in cancer risk assessment models.

Figure 4. Feature Importance for Cancer Risk Satisfaction

Accuracy and Precision

Accuracy measures the proportion of correctly identified high-risk patients, including (true positives (TP) and true negatives (TN)),relative to all diagnoses. It evaluates the overall validity of diagnostic model and itsefficacy in determining patient risk. The accuracy of the model is determinedby its ability to identify high-risk patients among all the instances it predicted as high-risk. Precision is computed as the ratio of TP to the sum of TP and FP.

Table 1. Overall outcome values for Accuracy and Precision

Methods	Accuracy (%)	Precision (%)
XGBoost [16]	84.5%	75%
SVM [16]	76.8%	60.2%
FlexiWPS-DBN[Proposed]	92.2%	91.5%

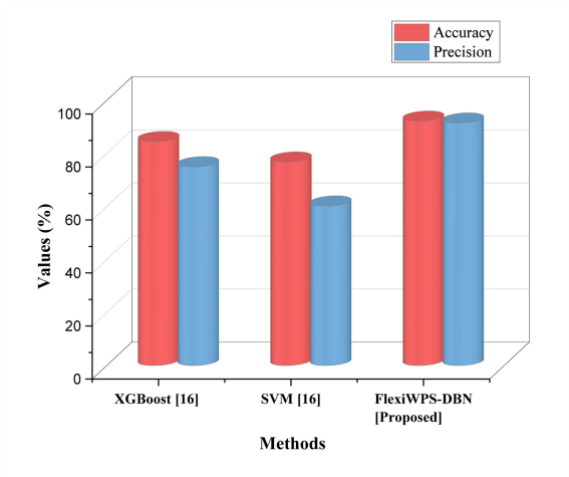

Figure 5. Accuracy and Precision Performance in Diagnostic Strategies

Accuracy and precision performance are presented in Figure 5 and Table 1. The existing methods, SVM and XGBoost model obtains76.8%&84.5%, respectively, whereas the suggested method, FlexiWPS-DBN achieves accuracy of 92.2%&precision of 91.5%. The FlexiWPS-DBN approach achieved better outcomes when compared to existing approaches in diagnostic strategies.

Recall & F1 Score

The model's recall, also known as sensitivity, evaluates how well it identifies all real high-risk patients. It is computed as the ratio of TP to the sum of FN and TP. The F1 score balances the two trades-offs

between precision and recall, determining the harmonic mean. The metric is particularly beneficial when there is an imbalance in high-risk patient classes.

Table 2. Overall performance for Recall and F1 score

Methods	Recall (%)	F1 score(%)
XGBoost [16]	75%	75%
SVM [16]	73.5%	66.1%
FlexiWPS-DBN[Proposed]	91.6%	90.3%

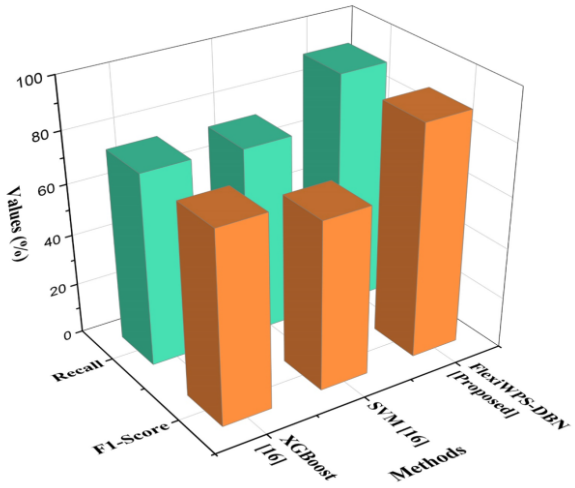

Figure 6. Recall and F1 Score Performance in Diagnostic Strategies

Recall andF1-score performance are presented in Figure 6 and Table 2. The existing methods, SVM and XGBoost model obtains recall of 73.5% & F1-score of 75%, respectively, whereas the suggested method, FlexiWPS-DBN achieves recall and f1-score is 91.6% & 90.3%. The FlexiWPS-DBN approach achieved better outcomes when compared to existing approaches in diagnostic strategies.

Discussion

FlexiWPS-DBN was presented earlier as a solution to challenges both SVM and XGBoost face, relying heavily on feature engineering, overfitting risks, and reduction of efficiency when dealing with complex, large-volume, or high-dimensional structured data. With FlexiWPS-DBN, these issues are completely addressed by migratively encompassing swarm intelligence alongside DL. FlexiWPS-DBN uses the automated selection and optimization of features, which makes the DL models more robust. Through a hierarchical quarternary structure, it captures all complex intricate regularities about high-risk patient

data, so it is of value for better diagnostics. Reducing the manual work and scalability would consequently almost always result in lower errors.

Conclusion

Predictive diagnostic strategies targeting identifying super high-risk patients rely on modern techniques of DL along with data-driven models. Clinico-demographic data are analyzed; the type of strategy is known to be exceptionally useful in furthering the early detection of targeted patients alongside improving the accuracy and efficacy of interventions in improving patient outcomes. All these methods provide the expenses related to the use of the laboratory parameter-driven approach, which ultimately supports patient-specific risk assessments and comprehensive population health management. Most laboratory parameters available in the market can be employed to create such helpful tools, as laying out the framework. Therefore, these fixed parameters have a significant impact on improving healthcare delivery and diagnostic accuracy. The technique was able to reliably and accurately stratify patients based on F1 score (90.3%), accuracy (92.2%), precision (91.5%), and recall (91.6%). FlexiWPS-DBN model has shown great performance in comparison with other modern diagnostic strategies. The proposed diagnostic strategy is governed by laboratory data availability and quality, which are not necessarily the same from one healthcare setting to another. Moreover, for model performance, diverse and representative datasets are essential to ensure generalizability. Future research can integrate real-time data of wearable devices for enhancing model adaptability to variable population and deploy in other chronic or rare diseases for more extensive healthcare usage.

References

1. I.M. Lubin, J.R. Astles, S. Shahangian, B. Madison, R. Parry, R.L. Schmidt, and M.L. Rubinstein, "Bringing the clinical laboratory into the strategy to advance diagnostic excellence," *Diagnosis*, vol. 8, no. 3, pp. 281-294, 2021. https://doi.org/10.1515/dx-2020-0119
2. A. Pulumati, A. Pulumati, B.S. Dwarakanath, A. Verma, and R.V. Papineni, "Technological advancements in cancer diagnostics: Improvements and limitations," *Cancer Reports*, vol. 6, no. 2, p. e1764, 2023. https://doi.org/10.1002/cnr2.1764
3. S. Kayikci and T.M. Khoshgoftaar, "Breast cancer prediction using gated attentive multimodal deep learning," *Journal of Big Data*, vol. 10, no. 1, p. 62, 2023. https://doi.org/10.1186/s40537-023-00749-w
4. J. Finkelstein, A. Smiley, C. Echeverria, and K. Mooney, "AI-driven prediction of symptom trajectories in cancer care: a deep learning approach for chemotherapy management," *Bioengineering*, vol. 11, no. 11, p. 1172, 2024. https://doi.org/10.3390/bioengineering11111172
5. J. Gong, J. Liu, H. Li, H. Zhu, T. Wang, T. Hu, M. Li, X. Xia, X. Hu, W. Peng, and S. Wang, "Deep learning-based stage-wise risk stratification for early lung adenocarcinoma in CT images: a multi-center study," *Cancers*, vol. 13, no. 13, p. 3300, 2021. https://doi.org/10.3390/cancers13133300
6. L. Zeng, L. Liu, D. Chen, H. Lu, Y. Xue, H. Bi, and W. Yang, "The Innovative Model Based on Artificial Intelligence Algorithms to Predict Recurrence Risk of Patients with Postoperative Breast Cancer," *Frontiers in Oncology*, vol. 13, p. 1117420, 2023. https://doi.org/10.3389/fonc.2023.1117420

7. J.C. Bortty, P.K. Bhowmik, S.A. Reza, I.A. Liza, M.N.I. Miah, M.S.R. Chowdhury, and M. Al Amin, "Optimizing lung cancer risk prediction with advanced machine learning algorithms and techniques," *Journal of Medical and Health Studies*, vol. 5, no. 4, pp. 35-48, 2024. https://doi.org/10.32996/jmhs
8. N.N. Phan, C.Y. Hsu, C.C. Huang, L.M. Tseng, and E.Y. Chuang, "Prediction of breast cancer recurrence using a deep convolutional neural network without region-of-interest labeling," *Frontiers in Oncology*, vol. 11, p. 734015, 2021. https://doi.org/10.3389/fonc.2021.734015
9. E. Briggs, M. de Kamps, W. Hamilton, O. Johnson, C.D. McInerney, and R.D. Neal, "Machine learning for risk prediction of oesophago-gastric cancer in primary care: comparison with existing risk-assessment tools," *Cancers*, vol. 14, no. 20, p. 5023, 2022. https://doi.org/10.3390/cancers14205023
10. A. Chennamadhavuni, L. Abushahin, N. Jin, C.J. Presley, and A. Manne, "Risk factors and biomarkers for immune-related adverse events: a practical guide to identifying high-risk patients and rechallenging immune checkpoint inhibitors," *Frontiers in Immunology*, vol. 13, p. 779691, 2022. https://doi.org/10.3389/fimmu.2022.779691
11. J. Cai, H. Chen, M. Lu, Y. Zhang, B. Lu, L. You, T. Zhang, M. Dai, and Y. Zhao, "Advances in the epidemiology of pancreatic cancer: Trends, risk factors, screening, and prognosis," *Cancer Letters*, vol. 520, pp. 1-11, 2021. https://doi.org/10.1016/j.canlet.2021.06.027
12. P. Khosravi, M. Lysandrou, M. Eljalby, Q. Li, E. Kazemi, P. Zisimopoulos, A. Sigaras, M. Brendel, J. Barnes, C. Ricketts, and D. Meleshko, "A deep learning approach to diagnostic classification of prostate cancer using pathology–radiology fusion," *Journal of Magnetic Resonance Imaging*, vol. 54, no. 2, pp. 462-471, 2021. https://doi.org/10.1002/jmri.27599
13. M.J. Iqbal, Z. Javed, H. Sadia, I.A. Qureshi, A. Irshad, R. Ahmed, K. Malik, S. Raza, A. Abbas, R. Pezzani, and J. Sharifi-Rad, "Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future," *Cancer Cell International*, vol. 21, no. 1, p. 270, 2021. https://doi.org/10.1186/s12935-021-01981-1
14. K.A. Tran, O. Kondrashova, A. Bradley, E.D. Williams, J.V. Pearson, and N. Waddell, "Deep learning in cancer diagnosis, prognosis, and treatment selection," *Genome Medicine*, vol. 13, p. 68, 2021. https://doi.org/10.1186/s13073-021-00968-x
15. Y. Xie, W.Y. Meng, R.Z. Li, Y.W. Wang, X. Qian, C. Chan, Z.F. Yu, X.X. Fan, H.D. Pan, C. Xie, and Q.B. Wu, "Early lung cancer diagnostic biomarker discovery by machine learning methods," *Translational Oncology*, vol. 14, no. 1, p. 100907, 2021. https://doi.org/10.1016/j.tranon.2020.100907
16. L. Meng, T. Wei, R. Fan, H. Su, J. Liu, L. Wang, X. Huang, Y. Qi, and X. Li, "Development and validation of a machine learning model to predict venous thromboembolism among hospitalized cancer patients," *Asia-Pacific Journal of Oncology Nursing*, vol. 9, no. 12, p. 100128, 2022. https://doi.org/10.1016/j.apjon.2022.100128