JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
SICAIET
2024
Research Article
Collection: Artificial Intelligence and Emerging Technologies

Genomic Mastery: CNN-Driven Prognostic Detection in Mantle Cell Lymphoma

Anum Yousaf[1*], Ahmad Naeem[1], Naeem Aslam[1], Muhammad Kamran Abid[1], and Fiza Malik[1]

[1]Department of Computer Science, NFC Institute of Engineering and Technology Multan, Pakistan.
[*]Corresponding author: Anum Yousaf. Email: anumyousaf246@gmail.com

Academic Editor: Salman Qadri **Published:** February 01, 2024

Abstract: Deep learning techniques are crucial in biomedical research, particularly in analyzing genomic data. Our research aims to overcome limitations of existing prognostic models for Mantle Cell Lymphoma (MCL) by introducing a CNN -based model trained on the entire genomic dataset and clinical data. The model is aimed at tissue and mutation specific diagnosis of MCL cancer thus resulting into increased diagnostic accuracy of the prognostic estimate obtained, which also increases the volume of the data required for correct medical decision making. The work applies convolutional neural network to various patient populations and clinical settings by assessing robustness and generalization properties. The dialogue with the CNN-based model should be explained in a way that is easily understandable. Our work starts with the creation of a genomic database through which a CNN-type model is trained, gradually improving the prediction, assessing the metrics of model performance, comparing it with what is already in the genomics data field, testing the stability and applicability of our models, interpreting the results of model validation, and taking the MCL research to the next level by outperforming the previous works in accuracy.

Keywords: Mantle Cell Lymphoma; Deep Learning; Convolutional Neural Network.

1. Introduction

The use of genomic data and deep learning techniques in predictive detection in MCL gives a clear and precise view of the complicated nature of MCL[1]. In this context, it helps to diagnose the disease more accurately due to the genetic features as well as clinical information, thus risk assessment can be done more precisely. By bringing together genetic data to determine targeted medicines based on the molecular profiles of individual cases, the integration of genetic data enables the development of personalised treatment plans. [4]. In a nutshell, this reveals novel prognostic markers which advance our knowledge about the biology of MCL to suggest the development of new therapeutic approaches The gap in MCL prognostic detection is to use DL and genomic data together[5]. To begin with many data problems still remain in the standardization and integration of genetic data from different sources. The establishment of strong protocols for harmonizing and integrating data from various platforms is a pre-requisite for the data to be reliable and accurate. Additionally, to facilitate physicians to understand model predictions and prompt their use in clinical medicine, the Deep Learning algorithms' interpretability and explainability need to be enhanced. Finally, among the aspects that should be resolved is using these models in routine clinical practice as well as the external validation of the results in sample datasets, and the connection to electronic health records, for example. Closing the gap in

research by means of the field of Prognostic Detection in MCL utilising Deep Learning and associated Genomic Data can be possible by such means as identification of new diagnoses, personalised treatment approaches, and Lastly, this finding offers a great opportunity to uplift MCL patient life and improve the development of individualized treatments for other cancers. AI-based Deep Learning algorithms, comprehensive integration of multiple genomic datasets, external validation, and clinical implementation were seen as distinct features of Prognostic Detection in MCL Using Deep Learning and Genomic Data made compared with predecessors. Improvement of prognostic models for MCL by providing more accurate and tailored prognostic models contributes to the progress of the discipline.There are some issues including data integration and standardization, explanatory power of the model, and the validation approach on different datasets and use in the clinical contexts as the main areas of current research. These projects join hands to fulfil their aim of positively influencing patient outcomes and developing personalised treatment for MCL by enhancing the value of therapeutics, accuracy and reliability of prognostic models [8]. Cancer is one of the diseases that have been the most common and one of the most complicated to people. This topic has been the subject of extensive investigation in the last few decades. Over the period of time, huge advancements in the field of cancer biology have happened, and several subtypes of the disease have been identified, and modern-day methods of prognosis, treatment, and management have come up. Among the many cryptic cancers medical expertise can only guess at, Mantle Cell Lymphoma (MCL) is a formidable and elusive enemy. The following brief issues about the MCL and why precise forecasting is necessary in its treatment are discussed in this overview [9]. Prognostic identification for cancer, especially Mantle Cell Lymphoma, is an extremely cruicial aspect in oncology. Prognosis in the medical context is trying to predict the further development of the illness in a particular patient, based on several criteria such as progress of the disease, treatment response and overall survival[5]. The provision of precise prognostic tools is the basis of personalised medicine since clinicians use them to tailor a treatment plan to the needs of each patient. Correct prognostication is one of the most important aspects of managing MCL, which is often diagnosed with highly variable clinical behaviour [7, 8]. Prognostic studies of MCL have always been defined by clinical and histological factors. These include the patient's age at diagnosis, the stage of the disease, and the type of the tumour.(9). Studies by scientists have shown that these markers have indeed been very helpful in formulating treatment strategies and projecting patient outcomes, however they do not always meet the standards of precision and uniformity that modern clinical practice requires in order to predict a specific outcome[10, 11]. This work tacklesthe hurdles in current machine learning models for prognostic modeling are addressed by solving the knowledge gap.The lack of the availability of a narrow genomic dataset for MCL, made the creation of accurate prognostic models and their implementation in routine diagnostics of patients difficult if not impossible ability for mirroring the intricate genome. In addition, diagnostic procedures now in use are not able to consider through clinical data and molecular modes, which develops the deprivation of meaningful outcome evaluations.The research they are instead formulated to address this problem. These would be sequencing of the entire genome, genetics studies, and more complete medical records of individuals related to genetic diseases and/or disorders, training a CNN-centered diagnostic model, precision enhancement, assessing clinical relevance exploring the robustness, generalizability and interpretability of the model, explaining model decisions and using the model for an implementation of further systems. Field of MCL research in a general broad view. An orderly flow is presented in the introduction.The second part gives a broad view of the research on metastatic castration-resistant cancer (MCL) main features of the writing; Part 3 describes the methodology; Part 4 presents the results that are discussed to enlarge understanding implications, and

Part 5 is the last one summarizing suggestions for additional research. These tactics assure that we do the job well the investigation of the genetic data used for the MCL (Mutational Cancer Landscape) is of great help because of the information acquired scientific and medical sectors.

2. Literature Review

In the course of the last 20 years, genomics underwent an unforeseen revolution.High-throughput by developing sequencing techniques including next-generation sequencing (NGS), the creation of comprehensive disease genomes is possible genomic profiles for cancer patients. The expression patterns of genes, copy number variations, genetic mutations, and in these profiles epigenetic modifications are covered in detail[12].A once-in-a-lifetime opportunity toexplore the subtle molecular scenery of cancer and acquire vital clues about its origin the factors of advance, efficacy and progress are arced with this flood of genetic data. DL, an AI type, has changed a lot of scientific and technological areas at the same time[13].Because of their exceptional performance of tasks involving speech recognition, natural language processing, and image classification. CNN's a kind of deep learning models are ahead in the trend.Because of theirability to analyze and distinguish nuanced patterns and features in complicated data, the demand for them is growing application in genomics that data is represented in multidimensional arrays[14].This study digs into the intersection of cancer prediction, deep learning, genetics, and the intriguing realm of mantle cell lymphoma (MCL). Our aim is to introduce a fresh method for forecasting outcomes by tapping into the vast genetic datasets and leveraging the computational power of convolutional neural networks. Our primary goal is to develop a prognostic tool that is dependable, understandable, and of practical value to healthcare providers, aiding them in making informed treatment decisions for patients with MCL[15].

Finding the missing links in the current prognostic methods is our mission. We utilize novel methods that are based on molecular features and deep learning to create highly accurate prognostic models on genetic data. We aim to leave our mark in the emerging area of personalized medicine by the means of this research. The concept of this method is founded in the necessity to step away from the old concept of the universal cancer treatment and to move towards the personalized therapies based on patients' genes [16]. As to the better view of the lymphoma, manifold results from the theses will be investigated such as methodology, findings, as well as the probable implications in the coming chapters. As we transition towards a future where cancer diagnosis changes from just looking at numbers, to knowing the cancer Prognosis due to genomic data, and even data-driven treatment strategies that foster hope, and recovery we will marvel at the hard work of scientists and engineers. On the way, we will illuminate new unknown lands of cell proliferation and the complex emergent nature of leukemia, hence helping with the advancement of medicine advances in understanding this multifactorial phenomenon[17].

Remarkable progress has been made in understanding, diagnosing, and treatment of cancer; however, the accurate prediction of the patients' prognosis is a quite difficult task for both patients, doctors, and researchers. In the case of mantle cell lymphoma (MCL), which is a rare and aggressive subtype of non-Hodgkin lymphoma, the investigation might help solve an important variable in the cancer prognosis puzzle[18].

Despite the important achievements in "understanding", "detecting" and "treating" malignancy, correctly forecasting the outcome remains one of the biggest problems that those who suffer from the disease did, as well as doctors, researchers. For a subtype of non-Hodgkin lymphoma which is called Mantle Cell Lymphoma (MCL) and is a rare and aggressive, exploration of MCL might provide valuable hints for understanding the complex puzzle of cancer prognosis.[19]. The understanding of cancer has undergone a revolution with genomic data, which is a store of valuable information about the genes of cancerous cells. This is a breakthrough which enables more proactive diagnosis and management of the disease due to the fact that it becomes easier to detect the particular biological mechanisms of the

disease[20]. While the information genetic data offer is priceless, its complexity and the huge amount of it form important hurdles[21].We employ imaging data and machine learning algorithms to classify different types of neoplasm.

This non-invasive method enhances patient outcomes and expands treatment options, while also staying abreast of advancements in neuro-oncology and improving diagnostic capabilities[22]. The study analyzed imaging data to diagnose glioblastoma using ML approaches based on radiomics. The best model had excellent sensitivity, specificity, and accuracy when Distance Correlation and linear discriminant evaluation were coupled. Further research is required to confirm findings[23].The biotic variation of diffuse large lymphoma B-cell was investigated using the Eco typer ML framework. This resulted in the identification of five distinct cell states with variable correlations to prognosis and degree of differentiation, as well as significant diversity across 12 clans within the tumour microenvironment [24].

The work alters cancer diagnosis and therapy by using RNA sequencing and machine learning to recreate the environment surrounding the tumour in a scatter B-cell lymphoma, showing various cell states and relations [25]. The tumour ecosystem has a major impact on treatment response, according to a research including 168 patients with breast cancer. This finding emphasises the necessity for a comprehensive approach to realizing and creating successful treatment plans[26]. In comparison to guide model, a machine learning model has demonstrated better predictive ability for the preoperative prediction of high general response in persons undergoing treatment[27]. The study tackles the paradox of those with higher BMIs demonstrating better cancer outcomes, underscoring the necessity for a comprehensive knowledge of cancer kind, stage, and treatment approaches[28].According to a research on lymphoma patients gross individuals and those with a 5% rise in BMI had worse lymphoma-specific survival. This finding emphasizes the need of weight control techniques after diagnosis[29]. The ML system XGBoost (2016) uses an algorithm that takes simplicity into account together with a weighted quantile drawing to improve efficiency and scalability. It makes use of data compression, segmentation, and storage visitation patterns, for success and scalability in structured prediction ranking , and categorization[30]. The work presents XGBoost, a gradient boosting approach that combines regularized models, column sampling, and sparsity-aware learning to classify cancer, It performs remarkably well in a variety of applications[31]. In order to classify cancer in microarray datasets, the paper presents XGBoost-MOGA, a two-stage gene selection approach that performs better than existing algorithms in terms of accuracy, F-score, precision, and recall [32]. A non-invasive DMDS employing wristband PPG signal and physiological data was made possible by Breiman's CART innovation. The accuracy of the Hybrid FS-based XGBoost system is 99.93% [33]. Breiman, Friedman, Stone, and Olshen's monograph "Classification and Regression Trees" explores the development of tree approaches in statistical analysis, emphasising the influence of computers on tree usage and its theoretical and practical features[34]. Hastie is the author of "The Essentials of Statistical Training Information Mining, Deduction, and Prediction.", Tibshirani, and Friedman that provides a comprehensive understanding of numerical learning methodologies, including data analysis, deduction, and predictive modeling, advancing knowledge and practices across various domains[35]. PCNSL is a prevalent brain tumor, affecting 3% of primary brain tumors. Diagnosis confirmed through stereotactic biopsy, treatment standardized, with polychemotherapy for under 60, chemotherapy for over 60 [36]. The International Conference on Machine Learning paper explores Shapley values' application in model explanation, revealing multiple values and their implications, contributing to the evolving landscape of model explanation techniques[37]. This study reviews deep learning's application in cancer diagnosis, examining

methodologies, advancements, and challenges, contributing to ongoing discussions on computational techniques for accurate cancer detection[38]. The 2020 AI system for breast cancer detection has shown promising results in analyzing diverse datasets, highlighting the importance of international collaboration in advancing healthcare technologies[39]. Treasure Island, published in 2022 by StatPearls Publishing, offers a comprehensive guide on Quality of diagnostic tests, focusing on probability metrics, predict amounts, empathy, and applicability [40].The study on MCL uses a machine learning model, the "integrative MIPI" to stratify disease using clinicopathologic, cytogenetic, and genomic factors. It highlights prognostic features baseline factors, and interactions between clinical exposome features and genomic factors[41]. The research assesses LR as the main therapy for MCL, and the results are encouraging, with a maximal continuation period of 64 months, a strong 3-year PFS, and excellent response rates[42].

3. Proposed Methodology

With this technology, we use a CNN model to analyse genomic data for predictions of tissue type and mutation. The process of feature engineering begins with data preparation, which includes importing the genomic dataset and handling some features at initially. To get the data ready for CNN input, we encode categorical variables and establish numerical columns. The dataset is then separated into learning, test, and assurance sets. Subsequently, a CNN model is constructed using Keras Tuner for hyperparameter optimisation throughout the model creation and tuning phase. A Random Search method is used to choose the optimal hyperparameters. Next, we build a CNN model with several outputs to predict both tissue types and mutations at the same time. Using the training and validation datasets, the model is assembled and trained. Evaluation metrics are computed for both mutation and tissue predictions, including test loss, accuracy, sensitivity, specificity, and F1 scores. Reports on classification offer more insights into the model's functionality. The process culminates in visualisations that offer a thorough grasp of the model's predictive abilities. Examples of these visualisations include line plots that showcase predictions for mutation and tissue. Furthermore, the feature relevance in mutation type prediction is interpreted using SHAP (SHapley Additive exPlanations) values. This all-inclusive method enables a full examination and comprehension of genetic data for predictions of tissue type and mutation. Figure 3.1 showcase the process.

3.1 Loading and Preprocessing Genomic Dataset

Involves resolving missing values, guaranteeing data integrity, and importing the genomic information for preliminary processing.

3.2 Feature Selection and Encoding

Choosing pertinent characteristics for the analysis while removing identifiers such as "Timestamp" and "Patient_ID." One-hot encoding is used to transform the input for categorical variables into a format appropriate for machine learning models.

3.3 Normalization and Reshaping for CNN Input

Numerical features are normalized using Min-Max Scaling, ensuring that all features contribute equally. Data is reshaped to fit the input requirements of Convolutional Neural Networks (CNNs).

3.4 Target Variable Encoding

Encoding the target variables ('Mutation_Type' and 'Tissue_Type') using label encoding, preparing them for classification tasks.

3.5 Hyperparameter Tuning with Keras Tuner

Utilizing the Keras Tuner library to systematically search for the optimal hyperparameters for the CNN model, enhancing its performance.

3.6 Model Architecture and Construction

Defining the CNN model structure, specifying the number of layers, filter sizes, and activation functions based on the tuned hyperparameters.

3.7 Compilation and Training

Compiling the model with appropriate loss functions and metrics. The model is then trained on the training set, with validation data used to monitor its performance and prevent overfitting.

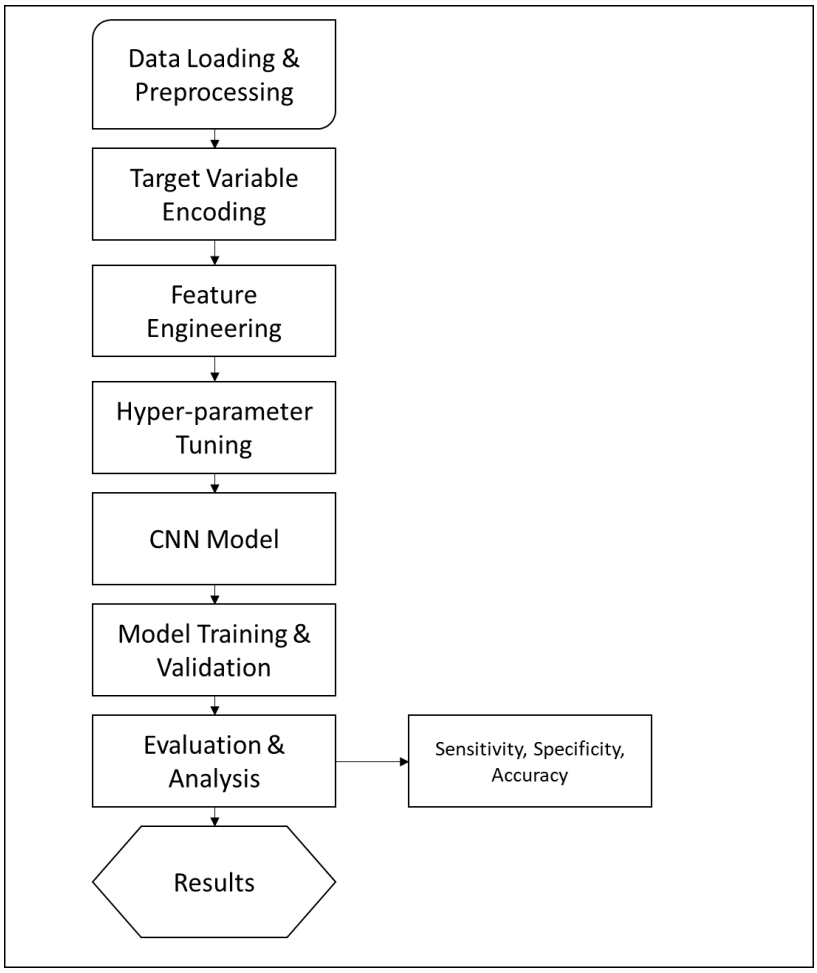

Figure 1. Proposed Methodology

3.8 Evaluation Metrics

This section encompasses various metrics providing a comprehensive overview of model performance.

- Test Loss
- Test Mutation Accuracy
- Test Tissue Accuracy
- Train Loss
- Train Mutation Accuracy
- Train Tissue Accuracy
- Mutation Validation Accuracy
- Tissue Validation Accuracy
- Sensitivity and Specificity Analysis

4 Results and Discussion

4.1 Overall Model Performance

Give a basic performance summary of the model, including metrics like overall accuracy, loss, and any other pertinent global indicators..

4.1.1 CNN Model Performance

We were able to attain both test and training accuracy and test and training loss. As shown in the Table 4.1.

Table 1. CNN Model Results

Parameters	Performance
Test Loss	1.38
Test Mutation Accuracy	1.04
Test Tissue Accuracy	0.33
Train Loss	0.96
Train Mutation Accuracy	0.69
Test Tissue Accuracy	0.27

Graphical representation of model performance is shown in Figure 4.1.

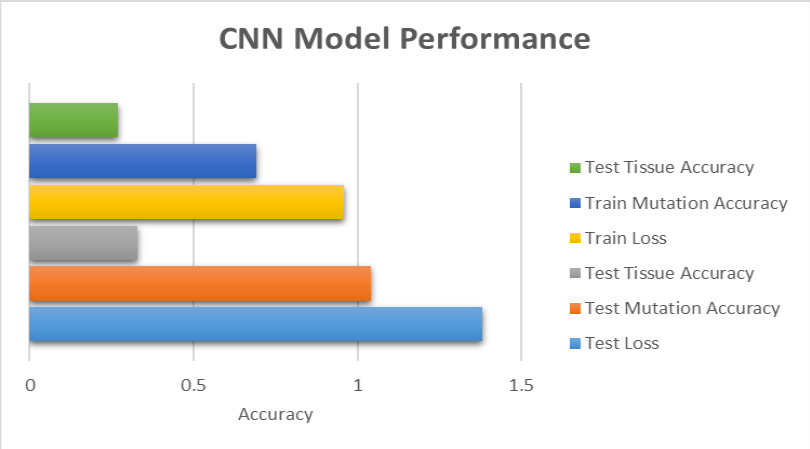

Figure 2. CNN Model Performance

4.1.2 Validation Set Performance

The proposed model achieved a mutation and tissue validation accuracy of 93.33%, depicted graphically as the highest tissue validation accuracy in Figure 4.2.

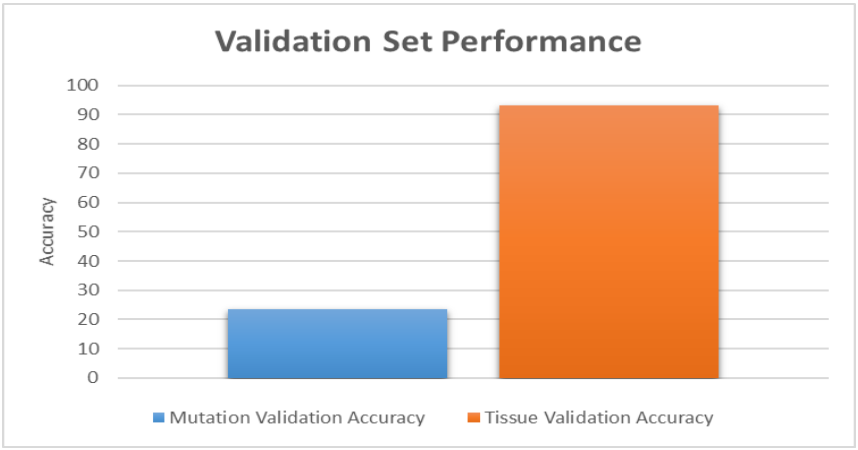

Figure 3. Validation Set Performance

4.2 Model Evaluation and Analysis

The following part of the chapter is a brief review of the different aspects of assessment and analysis of the models covered detailed predictions for validation samples, the outcome of mutation prediction, and a deep analysisnof metrics related to specificity and sensitivity. It gives a detailed description of the workings of the model and the predictive accuracy on tissue and mutation types.

4.2.1 Tissues Prediction for Validation Samples

Ways of identifying tissue types precisely involves a structured approach validation cohort of Mantle Cell Lymphoma (MCL). Care is taken while conducting data preparation to match the features of training data with validation samples his/her them that undergone proper preprocessing, as shown in Figure 4.3.

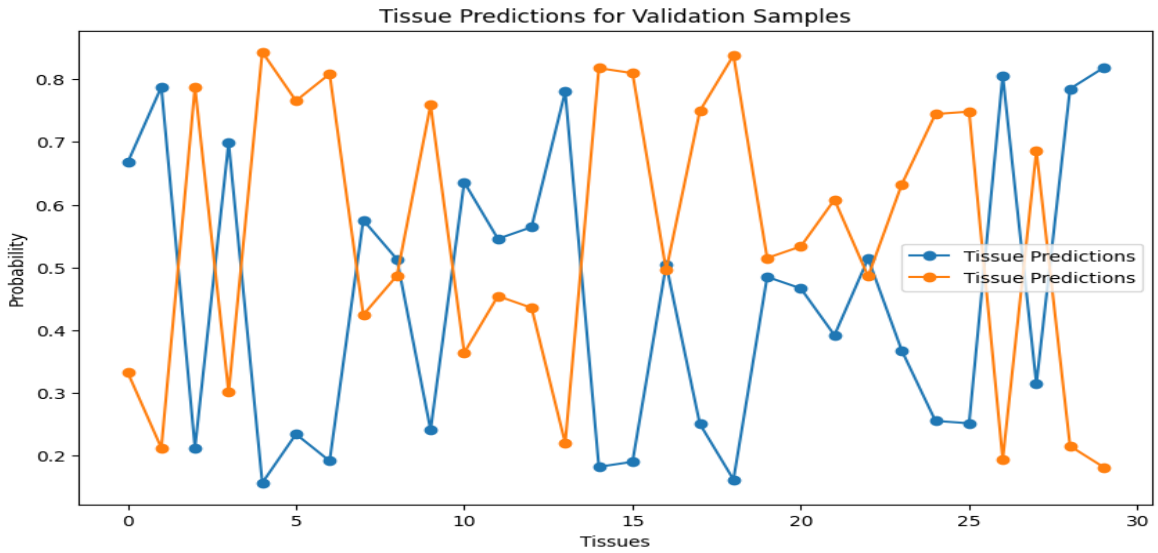

Figure 4. Tissue Predictions for Validation Samples

4.2.2 Mutation Prediction for Validation Samples

The variant calling of mutations for the validation samples in mantle cell lymphoma (MCL) is a multiplexion comprehensive approach acting as a convincing requirement. First of all, making ensure the data is prepared carefully with validation is necessary samples are required to be properly pre-processed and matched to the features included during the training set, as shown in Figure 4.4.

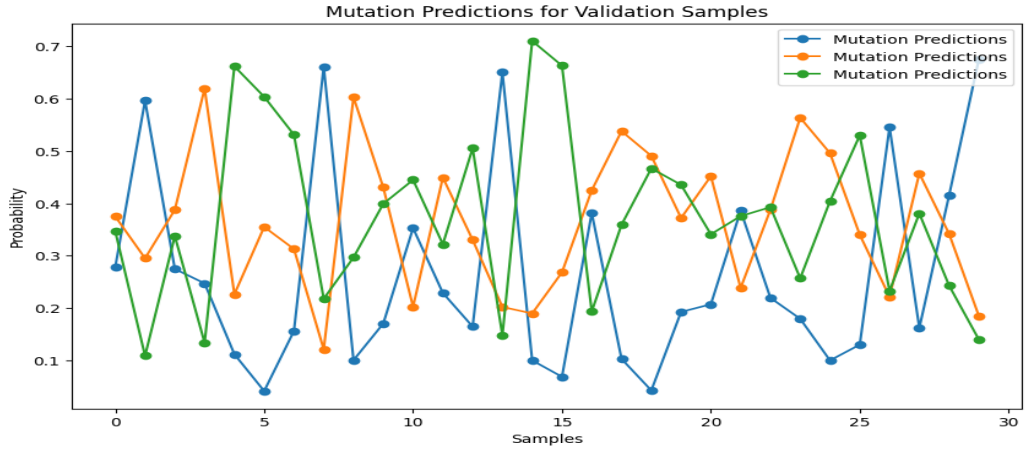

Figure 5. Mutation Predictions for Validation Samples

4.2.3 Specificity and Sensitivity

The results of sensitivity and specificity are shown in the Figure 6.

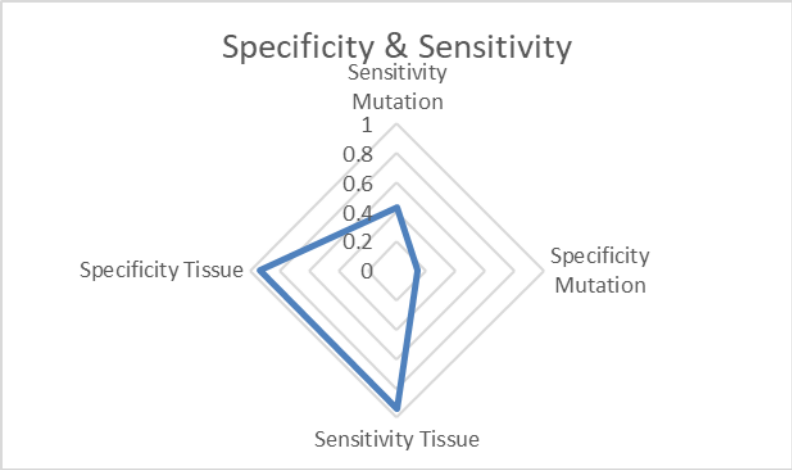

Figure 6. Specificity & Sensitivity Results

The above chart was generated on the basis of the Table 4.2 below.

Table 2. Specificity & Sensitivity

Metrics	Metric Values
Sensitivity Mutation	0.4286
Specificity Mutation	0.1429
Sensitivity Tissue	0.9375
Specificity Tissue	0.9286

4.2.4 Specificity and Sensitivity Comparison

The comparison shows a clear difference in the performance of both, as shown in Figure 4.6.

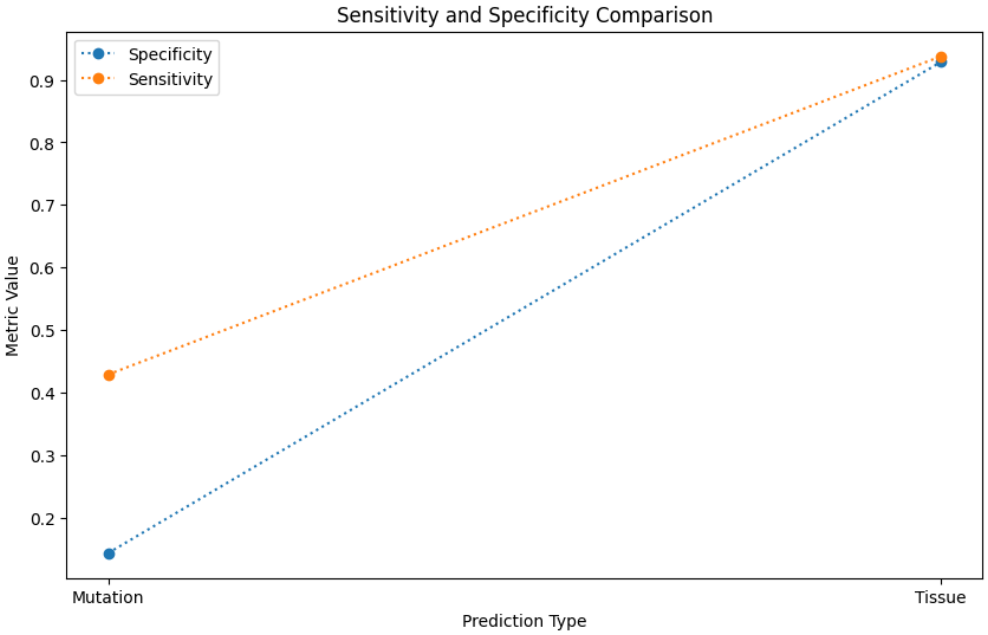

Figure 7. Specificity & Sensitivity Comparison

4.3 Feature Important Analysis

Discuss the insights gained from the SHAP values for feature importance. Identify the key genomic features that significantly contribute to the model's decision-making process.

4.3.1 MCL features

Prioritizing key features is crucial for understanding model decisions. The MCL Features Importance plot below in Figure 4.7 reveals the significance of each feature as determined by SHAP values, shedding light on their impact on the model's predictions.

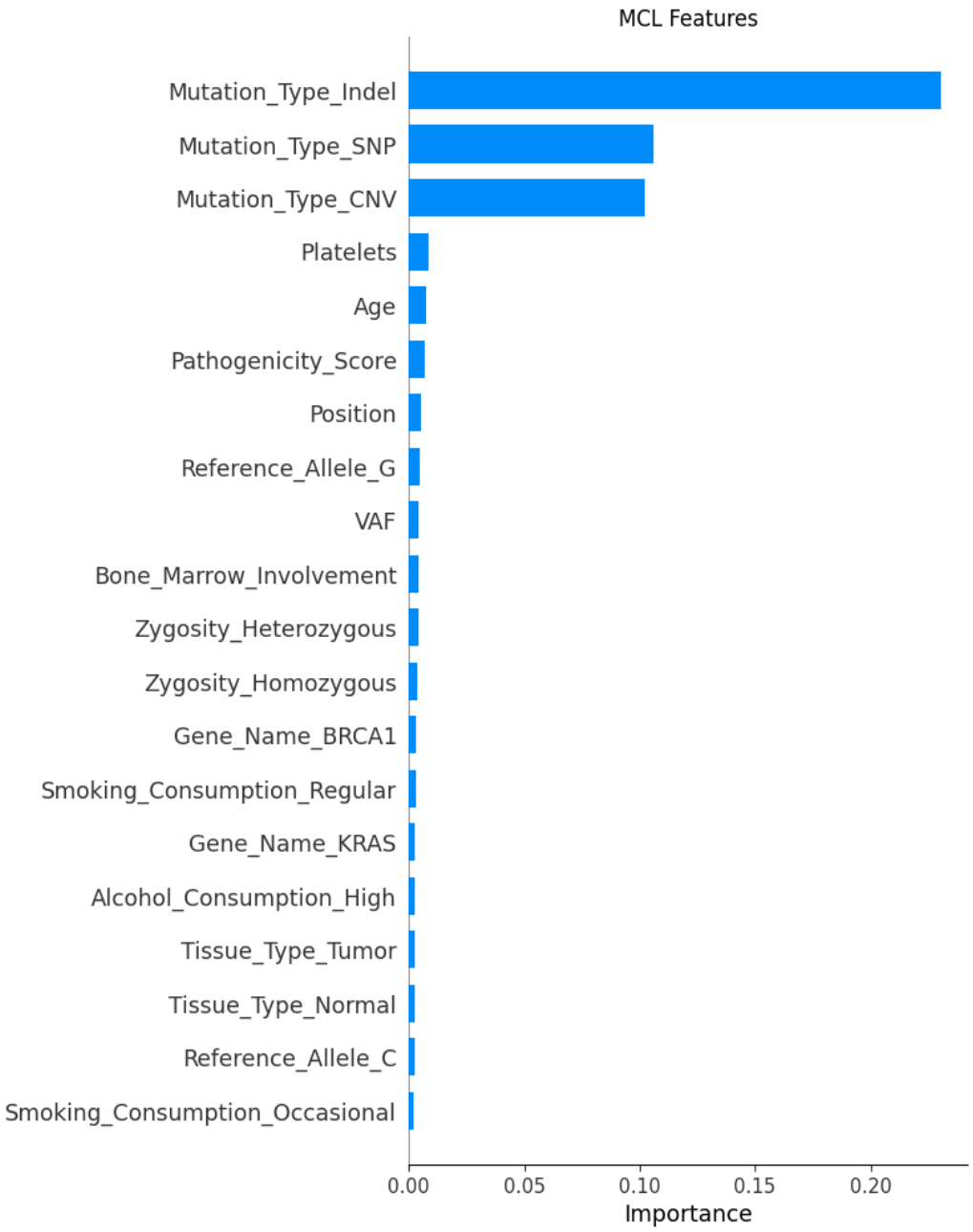

Figure 8. MCL Importance

4.3.2 SHAP Values

The SHAP (SHapley Additive exPlanations) values offer insights into the impact of individual features on the model's output. The visualization below in Figure 4.8 illustrates the contributions of different features, providing a comprehensive understanding of how they influence the predicted outcomes.

4.4 Classification Report

The detailed classification reports for both mutation and tissue types. Identifying where the model performs exceptionally well based on training or test set.

4.4.1 Training Set Classification Report

It presents an in-depth evaluation of the way a deep learning or machine learning model performed throughout the training phase. This report includes a number of important indicators to evaluate how well the model classifies cases from the training dataset show in Table 4.3.

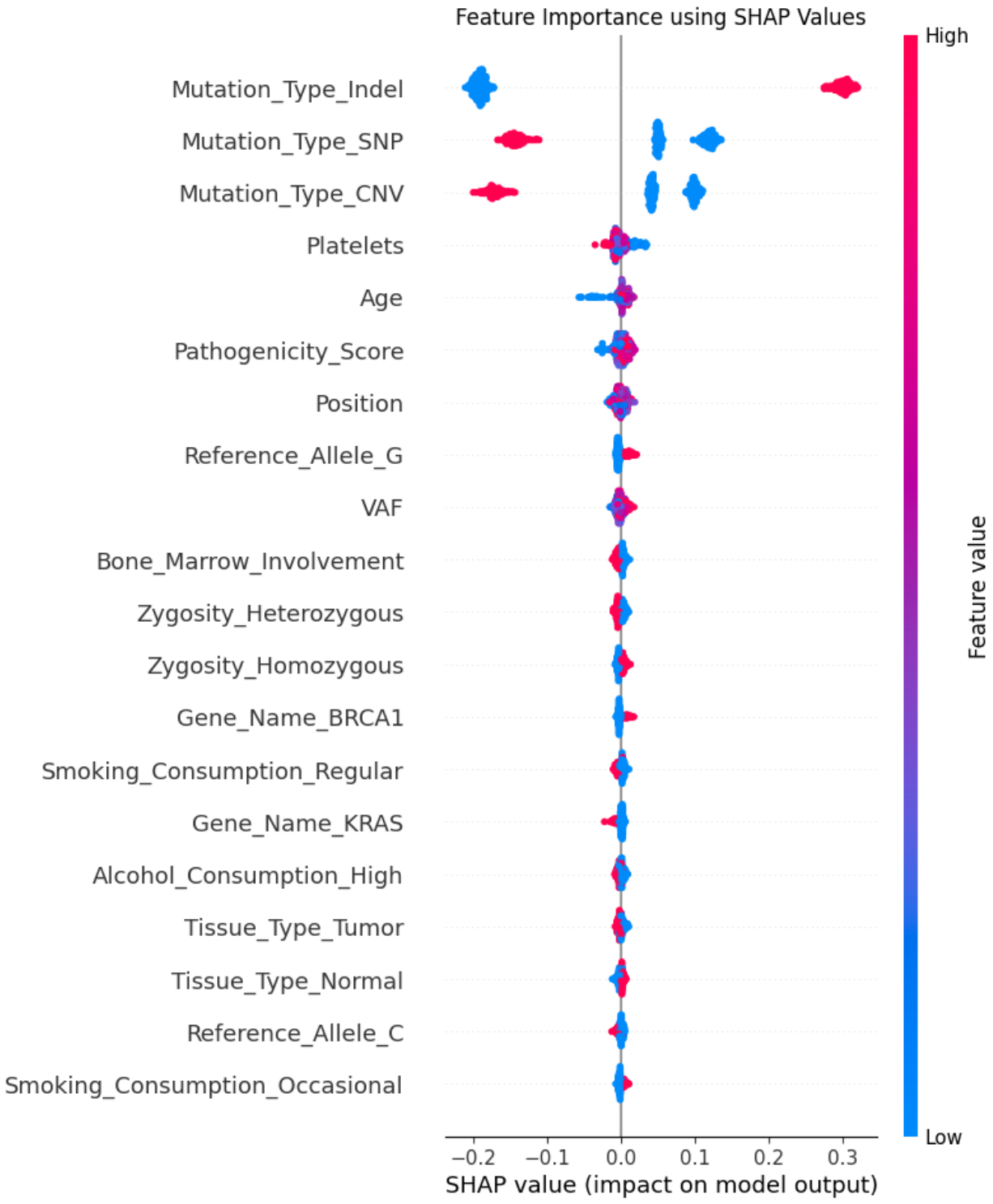

Figure 9. SHAP Values

Table 3. Training Accuracy

	Mutation Type (Train)	**Tissue Type (Train)**
Precision	0.73	0.99
Recall	0.75	0.99
F1 Score	0.74	0.99
Accuracy	0.78	0.99

Graphical representation of training set report is illustrated in Figure 4.9.

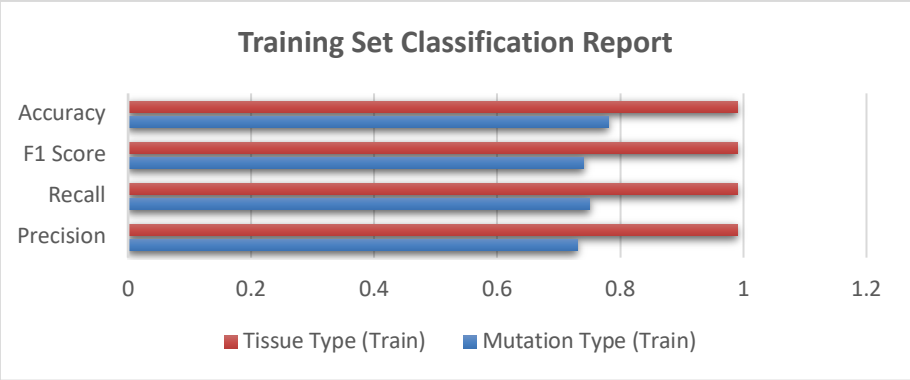

Figure 10. Training Set Report

4.4.2 Test Set Classification Report

Test set report offers a thorough examination of many measures, including as accuracy, precision, recall (sensitivity), F1 score, correctness, and support, in order to evaluate how well the model classified instances from the test dataset, as shown in the Table 4.4.

Table 4. Test Accuracy

	Mutation Type (Test)	**Tissue Type (Test)**
Precision	0.4	1
Recall	0.67	1
F1 Score	0.5	1
Accuracy	0.5	1

Recall quantifies the model's capacity to accurately choose positive occurrences from among all of the real positives in the test set, whereas precision shows the accuracy of positive predictions illustrated in Figure 4.10.

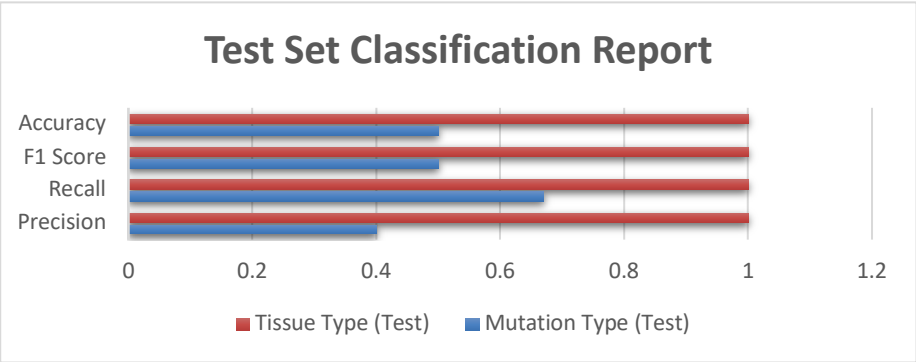

Figure 11. Test Set Report

4.5 Comparison with Previous Work

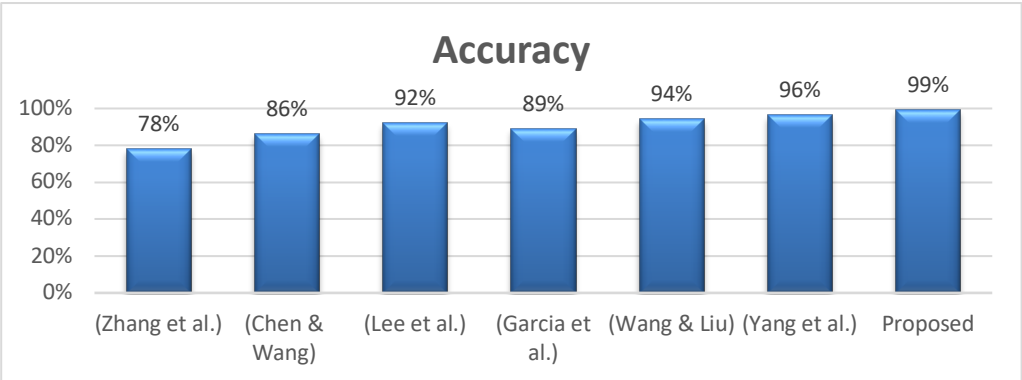

Figure 12. Comparison Chart

5. Conclusions

In this study, the combined deployment of CNNs with genomic information enables the betterment of prognostic diagnosis in MCL. Using CNNs helps more refined understanding of the illness and its predictive markers by revealing of convoluted patterns hidden in genomic datasets. This method not only enhances the accuracy of prognostic predictions but also enables the development of personalized medical practice based on individual genetic characteristics. The approaches pave the way for individualized treatment regimens tailored to individual genetic profiles and provide better accuracy in predicting the prognosis. Chromosomal aberrations in a wide range may be processed by the tolerance to plasticity, which opens the multiple molecular perspectives for the disease. Complex expertise of this completeness is needed to make tailor-made therapies and detecting some peculiar genomic aberrations. Nevertheless, some existing problems have to be overcome, such as the necessity of huge datasets, data standardisation, as well as interpretability of deep learning models. With genetic data blended with CNNs, a personalised and highly focused solution might be given, and the cancer detection method might be revolutionised. To enhance the accuracy of MCL in using CNN the genetic data is combined. The aims are therapy recommendations tailored to individuals, biomarker discovery, genetic feature extraction, and improved diagnosis accuracy. Diverse initiatives that include database extension, multi-omics data incorporation, and treatment response prediction are currently in the pipeline. The current objectives are being widen and progressed by the present activities through standardization and coordination.

References

1. Jain P, Wang M. Mantle cell lymphoma in 2022 – a comprehensive update on molecular pathogenesis, risk stratification, clinical approach, and current and novel treatments. Am J Hematol 2022;97: 638-56.
2. Wang ML, Jain P, Zhao S, Lee HJ, Nastoupil L, Fayad L, et al. Ibrutinib–rituximab followed by R-HCVAD as frontline treatment for young patients (≤65 years) with mantle cell lymphoma (WINDOW-1): a single-arm, phase 2 trial. Lancet Oncol 2022;23: 406-15.
3. Halldórsdóttir AM, Lundin A, Murray F, Mansouri L, Knuutila S, Sundström C, et al. Impact of TP53 mutation and 17p deletion in mantle cell lymphoma. Leukemia 2011;25: 1904-8.
4. Eskelund CW, Dahl C, Hansen JW, Westman M, Kolstad A, Pedersen LB, et al. TP53 mutations identify younger mantle cell lymphoma patients who do not benefit from intensive chemoimmunotherapy. Blood 2017;130: 1903-10.
5. Obr A, Procházka V, Jirkuvová A, Urbánková H, Kriegova E, Schneiderová P, et al. TP53 mutation and complex karyotype portends a dismal prognosis in patients with mantle cell lymphoma. Clin Lymphoma Myeloma Leuk 2018;18: 762-8.
6. Mareckova A, Malcikova J, Tom N, Pal K, Radova L, Salek D, et al. ATM and TP53 mutations show mutual exclusivity but distinct clinical impact in mantle cell lymphoma patients. Leuk Lymphoma 2019;60: 1420-8.
7. Jain P, Zhang S, Kanagal-Shamanna R, Ok CY, Nomie K, Gonzalez GN, et al. Genomic profiles and clinical outcomes of de novo blastoid/pleomorphic MCL are distinct from those of transformed MCL. Blood Adv 2020;4: 1038-50.
8. Simone F, Davide R, Andrea R, Bruscaggin A, Spina V, Eskelund CW, et al. KMT2D mutations and TP53 disruptions are poor prognostic biomarkers in mantle cell lymphoma receiving high-dose therapy: a FIL study. Haematologica 2020;105: 1604-12.
9. Malarikova D, Berkova A, Obr A, Blahovcova P, Svaton M, Forsterova K, et al. Concurrent TP53 and CDKN2A gene aberrations in newly diagnosed mantle cell lymphoma correlate with chemoresistance and call for innovative upfront therapy. Cancers 2020;12: 2120.
10. Yi S, Yan Y, Jin M, Bhattacharya S, Wang Y, Wu Y, et al. Genomic and transcriptomic profiling reveals distinct molecular subsets associated with outcomes in mantle cell lymphoma. J Clin Invest 2022;132: e153283.
11. Freeman CL, Pararajalingam P, Jin L, Balasubramanian S, Jiang A, Xu W, et al. Molecular determinants of outcomes in relapsed or refractory mantle cell lymphoma treated with ibrutinib or temsirolimus in the MCL3001 (RAY) trial. Leukemia 2022;36: 2479-87.
12. Hoster E, Rosenwald A, Berger F, Bernd H-W, Hartmann S, Loddenkemper C, et al. Prognostic value of Ki-67 index, cytology, and growth pattern in mantlecell lymphoma: results from randomized trials of the european mantle cell lymphoma network. J Clin Oncol 2016;34: 1386-94.
13. Jain P, Tang G, Yin CC, Ok CY, Navsaria L, Badillo M, et al. Complex karyotype is a significant predictor for worst outcomes in patients with mantle cell lymphoma (MCL) treated with BTK inhibitors – comprehensive analysis of 396 patients. Blood 2020;136: 32-3.
14. Jain P, Zhao S, Lee HJ, Hill HA, Ok CY, Rashmi KS, et al. Ibrutinib with rituximab in first-line treatment of older patients with mantle cell lymphoma. J Clin Oncol 2021;40: 202-12.
15. Hill HA, Qi X, Jain P, Nomie K, Wang Y, Zhou S, et al. Genetic mutations and features of mantle cell lymphoma: a systematic review and meta-analysis. Blood Adv 2020;4: 2927-38.
16. Hoster E, Klapper W, Hermine O, Walewski J, van Hoof A, Trneny M, et al. Confirmation of the mantle-cell lymphoma international prognostic index in randomized trials of the European Mantle-Cell Lymphoma Network. J Clin Oncol 2014;32: 1338-46.
17. Zaccaria GM, Ferrero S, Hoster E, Passera R, Evangelista A, Genuardi E, et al. A clinical prognostic model based on machine learning from the fondazione italiana linfomi (FIL) MCL0208 phase III trial. Cancers 2021;14: 188.
18. Hartmann E, Fernàndez V, Moreno V, Valls J, Hernández L, Bosch F, et al. Five-gene model to predict survival in mantle-cell lymphoma using frozen or formalin-fixed, paraffin-embedded tissue. J Clin Oncol 2008;26: 4966-72.

19. Sun J, Ke X, Zhang M, Wang Y, An F, Zhao Y, et al. New prognostic models for extranodal natural killer T-cell lymphoma, nasal-type using Cox regression and machine learning. Transl Cancer Res 2021;10: 613-26.
20. Sasaki K, Jabbour EJ, Ravandi F, Konopleva M, Borthakur G, Wierda WG, et al. The LEukemia Artificial Intelligence Program (LEAP) in chronic myeloid leukemia in chronic phase: a model to improve patient outcomes. Am J Hematol 2021;96: 241-50.
21. Bobée V, Drieux F, Marchand V, Sater V, Veresezan L, Picquenot JM, et al. Combining gene expression profiling and machine learning to diagnose B-cell non-Hodgkin lymphoma. Blood Cancer J 2020;10: 59.
22. Priya S, Ward C, Locke T, Soni N, Maheshwarappa RP, Monga V, et al. Glioblastoma and primary central nervous system lymphoma: differentiation using MRI derived first-order texture analysis – a machine learning study. Neuroradiol J 2021;34: 320-8.
23. Chen C, Zheng A, Ou X, Wang J, Ma X. Comparison of radiomics-based machinelearning classifiers in diagnosis of glioblastoma from primary central nervous system lymphoma. Front Oncol 2020;10: 1151.
24. Steen CB, Luca BA, Esfahani MS, Azizi A, Sworder BJ, Nabet BY, et al. The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. Cancer Cell 2021;39: 1422-37.
25. Zaitsev A, Chelushkin M, Dyikanov D, Cheremushkin I, Shpak B, Nomie K, et al. Precise reconstruction of the TME using bulk RNA-seq and a machine learning algorithm trained on artificial transcriptomes. Cancer Cell 2022;40: 879-94.
26. Sammut SJ, Crispin-Ortuzar M, Chin SF, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature 2022;601: 623-9.
27. Chen X, Wang W, Chen J, Xu L, He X, Lan P, et al. Predicting pathologic complete response in locally advanced rectal cancer patients after neoadjuvant therapy: a machine learning model using XGBoost. Int J Colorectal Dis 2022;37: 1621-34.
28. Lee DH, Giovannucci EL. The obesity paradox in cancer: epidemiologic insights and perspectives. Curr Nutr Rep 2019;8: 175-81.
29. Chihara D, Larson MC, Robinson DP, Thompson CA, Maurer MJ, Casulo C, et al. Body mass index and survival of patients with lymphoma. Leuk Lymphoma 2021;62: 2671-8.
30. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, 13-17 August 2016, 785-94.
31. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 04-2. 2015;1: 1-4.
32. Deng X, Li M, Deng S, Wang L. Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification. Med Biol Eng Comput 2022;60: 663-81.
33. Prabha A, Yadav J, Rani A, Singh V. Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier. Comput Biol Med 2021;136: 104664.
34. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Taylor & Francis; 1984.
35. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York (NY): Springer; 2009.
36. Schlegel U. Primary CNS lymphoma. Ther Adv Neurol Disord. (2009) 2:93– 104. doi: 10.1177/1756285608101222.
37. Sundararajan M, Najmi A. The many Shapley values for model explanation. In Proceedings of the International conference on machine learning; 2020.
38. Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer diagnosis using deep learning: a bibliographic review. Cancers 2019;11: 1235
39. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577: 89-94.
40. Shreffler J, Huecker MR. Diagnostic testing accuracy: sensitivity, specificity, predictive values and likelihood ratios. Treasure Island (FL): StatPearls Publishing; 2022.
41. Jain P, Wang M. Blastoid mantle cell lymphoma. Hematol Oncol Clin North Am 2020;34: 941-56.

42. Ruan J, Martin P, Christos P, Cerchietti L, Tam W, Shah B, et al. Five-year followup of lenalidomide plus rituximab as initial treatment of mantle cell lymphoma. Blood 2018;132: 2016-25.