IEEE Journal of Translational Engineering in Health and Medicine

Received 18 September 2023; revised 22 January 2024; accepted 27 January 2024. Date of publication 31 January 2024; date of current version 14 February 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3360865

Deep Learning-Based, Multiclass Approach to Cancer Classification on Liquid Biopsy Data

MAKSYM A. JOPEK[1,2], KRZYSZTOF PASTUSZAK[2,3], SEBASTIAN CYGERT[4,5], MYRON G. BEST[6], THOMAS WÜRDINGER[6], JACEK JASSEM[7], ANNA J. ŻACZEK[1], AND ANNA SUPERNAT[1,2]

[1]Laboratory of Translational Oncology, Intercollegiate Faculty of Biotechnology, University of Gdańsk, 80-309 Gdańsk, Poland
[2]Centre of Biostatistics and Bioinformatics, Medical University of Gdańsk, 80-210 Gdańsk, Poland
[3]Department of Algorithms and Systems Modeling, Telecommunications and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
[4]Department of Multimedia Systems, Faculty of Electronics, Telecommunications, and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
[5]Ideas NCBR, 00-801 Warsaw, Poland
[6]Department of Neurosurgery, Amsterdam UMC Locatie AMC, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
[7]Department of Oncology and Radiotherapy, Medical University of Gdańsk, 80-210 Gdańsk, Poland

CORRESPONDING AUTHOR: A. SUPERNAT (anna.supernat@gumed.edu.pl)

This work was supported in part by the SONATA Grant of the National Science Centre under Grant 2018/31/D/NZ5/01263, in part by the Medical University of Gdańsk Statutory Work under Grant ST-23 and Grant 02-0023/07, and in part by the National Center for Research and Development through the Lider XI under Grant 0059/L-11/2019.
This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Independent Ethics Committee of the Medical University of Gdansk under Application No. NKBBN/434/2017, and performed in line with the Helsinki Declaration, as revised in 1983.
This article has supplementary downloadable material available at https://doi.org/10.1109/JTEHM.2024.3360865, provided by the authors.

ABSTRACT The field of cancer diagnostics has been revolutionized by liquid biopsies, which offer a bridge between laboratory research and clinical settings. These tests are less invasive than traditional biopsies and more convenient than routine imaging methods. Liquid biopsies allow studying of tumor-derived markers in bodily fluids, enabling the development of more precise cancer diagnostic tests for screening, disease monitoring, and therapy personalization. This study presents a multiclass approach based on deep learning to analyze and classify diseases based on blood platelet RNA. Its primary objective is to enhance cancer-type diagnosis in clinical settings by leveraging the power of deep learning combined with high-throughput sequencing of liquid biopsy. Ultimately, the study demonstrates the potential of this approach to accurately identify the patient's type of cancer. *Methods:* The developed method classifies patients using heatmap images, generated based on gene expression arranged according to the Kyoto Encyclopedia of Genes and Genomes pathways. The images represent samples of patients with ovarian cancer, endometrial cancer, glioblastoma, non-small cell lung cancer, and sarcoma, as well as cancer patients with brain metastasis. *Results:* Our deep learning-based models reached 66.51% balanced accuracy when distinguishing between those 6 sites of cancer origin and 90.5% balanced accuracy on a location-specific dataset where cancer types from close locations were grouped. The developed models were further investigated with an explainable artificial intelligence-based approach (XAI) - SHAP. They returned a set of 60 genes with the highest impact on the models' decision-making process. *Conclusions:* Our results show that deep-learning methods are a promising opportunity for cancer detection and could support clinicians' decision-making process in finding the solution for the black-box problem.

INDEX TERMS Deep learning, liquid biopsies, tumor educated platelets (TEP), CNN-convolutional neural network, explainable AI.

Clinical and Translational Impact Statement— Utilizing TEPs-based liquid biopsies and deep learning, our study offers a novel approach to early cancer detection, highlighting cancer origin. The integration of Explainable AI reinforces trust in predictive outcomes. Category: Early/Pre-Clinical Research.

I. INTRODUCTION

THE number of patients diagnosed with cancer continues to rise each year due to carcinogen exposure, higher life expectancy, health awareness in society, and access to diagnostic tests [1]. Routine markers such as carbohydrate antigen 125 (CA-125) [2] in ovarian cancer or

IEEE Journal of Translational Engineering in Health and Medicine

carcinoembryonic antigen (CEA) in non-small cell lung cancer (NSCLC) [3] and imaging methods are successfully used in diagnosis only to some extent. Routinely used diagnostic tests are often costly, hard to obtain, or lack accuracy. Hence, a non-invasive, straightforward approach is needed to enable an extended blood morphology that detects cancer and indicates its type.

This problem can be partially addressed by introducing liquid biopsies – samples of body fluids, especially blood, analyzed with the latest high-throughput technologies [4]. One sample can be analyzed in multiple ways, including interrogation of extracellular vesicles, circulating proteins, cell-free RNA, cell-free DNA, RNA of tumor-educated platelets (TEPs), or RNA/DNA of single circulating tumor cells (CTCs), immune cells, circulating endothelial cells and cancer-associated fibroblasts [4], [5]. There are many breakthrough cancer studies related to analyzing those biomarkers where each type may shine in its field, covering the following areas: a) screening and early cancer detection research (EpiproColon, GRAIL) [6], [7]; b) tools for decision-making and companion diagnostics (Cobas EGFR Mutation Test v2, Guardant360) [8], [9]; as well as c) determination whether the patient has a drug targetable mutation (FoundationOne-Liquid, Resolution-HRD) [10]. Multi-omics models and multi-platform solutions have also emerged, basing their outcomes on data from multiple related sources [11], [12], [13], [14]. In general, liquid biopsies constitute a promising alternative to traditionally performed tissue biopsies and, in certain conditions, can complement, if not replace, traditional biopsy evaluation [15], [16].

TEPs have shown potential in early cancer detection, tracking cancer's evolution, and metastasis [17], [18]. They can be used complementarily with other biomarkers like ctDNA in the detection of endometrial cancer [16] and along the CA-125 to boost the detection of ovarian cancer [19]. Impressively, platelet transcriptomes are fairly stable, retaining their RNA profiles even under prolonged storage at room temperature or freezing, making them practical study subjects [20]. Past research indicates that RNA profiles from TEPs can provide diagnostic accuracy of up to 80% for detecting sites of origin of various cancers [17], [21].

In our work, we harness the power of deep learning (DL) models, building upon the approach by [18] organizing genes within signaling pathways according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) [22]. Our analysis involves a thorough comparison of ResNet, DenseNet, and EfficientNet models. Additionally, to ensure our models are both accurate and transparent, we've employed an Explainable AI (XAI) technique, SHAP [23]. This method illuminates the inner workings of our models, helping us understand how specific predictions are made and the significance of different genes in the outcome.

To sum up, the contributions of this work are:

- Deep Learning with TEPs: We've utilized TEPs-based liquid biopsies and advanced deep learning for early cancer detection, presenting a potential tool to help medical professionals pinpoint cancer origins.
- Benchmarking: Our model, benchmarked against prominent studies, offers a fresh perspective on pan-cancer classification challenges and opportunities.
- Integration of XAI: By incorporating SHAP, we've deepened our understanding of cancer's molecular markers and reinforced trust in our model's predictive capabilities.

TABLE 1. Data source overview.

Disease status	Number of samples			Data source
	Quality passed	Quality rejected	Rejected due to confounding factors	
Asymptomatic donor (AD)	338	0	0	GSE156902 [25]
Ovarian cancer (OC)	28	0	0	GSE158508 [18]
Endometrial cancer (EC)	37	2	0	GSE184904 [16]
Non-small cell lung cancer (NSCLC)	397	4	0	GSE89843 [26]
Glioblastoma (GBM)	84	0	150	GSE156902 [25]
Brain metastasis (BM)	36	0	19	GSE156902 [25]
Sarcoma (SARC)	38	24	0	[27]

II. METHODS AND PROCEDURES

Platelet RNA samples were processed according to the guidelines published by [24]. All patients signed informed consent. The study was performed following the Helsinki Declaration, as revised in 1983. The final dataset contained the data from asymptomatic donors (AD) and patients with 5 different cancer types (NSCLC – non-small cell lung cancer, OC – ovarian cancer, EC – endometrial cancer, GBM – glioblastoma, SARC – sarcoma) and brain metastasis (BM). Patients with NSCLC and BM were treated as the NSCLC group, though patients with brain metastases also had other primary tumors than NSCLC. The blood samples from patients enrolled to the study were collected before the start of the treatment. Hence, there were no confounding factors related to the prior cancer therapy. Due to quality control, we removed all samples with less than 100k total reads from the dataset and only used genes with known Gencode status. All samples underwent the same pre-processing and were aggregated and normalized together. Table 1 summarizes all elements of the dataset.

Following the methods of [18], expression data normalization was performed via the DESeq2 package in R [28] using Variance Stabilizing Transformation [29]. The Gencode v19 GRCh37 served as the annotation tool. Meeting quality checks, expression profiles were transformed into visual formats, each pixel mirroring the expression level of a distinct gene, akin to a heatmap. Organizing these genes based on the KEGG pathways [22] via the GAGE package [30] yielded 345 $\times$ 243 expression arrays, where each row signified a

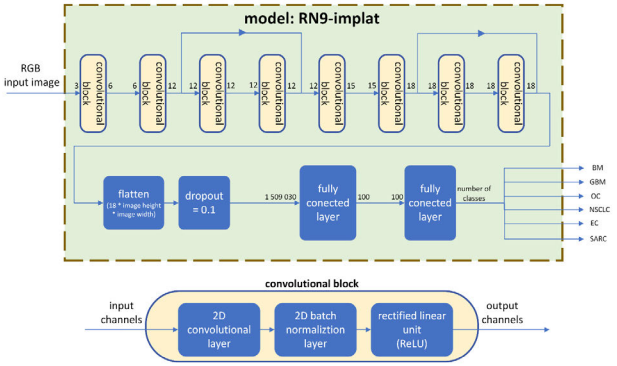

FIGURE 1. Model RN9-implat based on RN9 backbone. It comprises eight convolutional blocks for feature extraction and two fully connected layers responsible for classification tasks. A dropout layer was added to prevent overfitting. Numbers depict the amount of input and output channels for each layer.

specific pathway associated with metabolism, genetic information processing, environmental information processing, cellular processes, immune system, other organismal systems, cancer, or other human diseases. While the order of genes within the row mattered, the impact on the performance of the order of the rows was minimal as it was shown in one of our recent studies [31].

Our objective was to assess various deep-learning frameworks for cancer classification through liquid biopsy data. We evaluated the architectural depth and intricacy's influence on the models' accuracy. Initial experiments utilized ResNet architectures, recognized for computational efficiency [31]. Three primary variations, including ResNet9, ResNet18, and ResNet50, were assessed with and without MaxPooling layers, hypothesizing the potential loss of vital gene information with pooling at two different kernel sizes. In addition, we conducted an extended comparative analysis with other well-known architectures, including DenseNet-121, DenseNet-169, EfficientNet-b0, and EfficientNet-b4. Interestingly, our comprehensive evaluation revealed that the more complex backbones of DenseNet and EfficientNet did not enhance performance in our specific application. The shallower architectures, including RN9 and RN18, consistently outperformed other models. As a result, our efforts concentrated on refining the ResNet9's configurations through hyperparameter tuning and kernel optimization. The eventual model, termed RN9-implat, discarded pooling layers, expanding its size but enhancing its accuracy. The complete model architecture is shown in Figure 1, and detailed classification results are available in Supplementary Material ("Appendix.1 – model metrics").

Developed in Python and reliant on the PyTorch backend [32], our models were trained using cross-entropy as the loss function, with gradient optimization achieved through the ADAM algorithm [33]. Various random seeds and learning ratios of 0.01 and 0.001 were taken into account. Balanced accuracy (calculated as the average recall of each class) served as the main performance metric. To further evaluate the performance of models we measured the accuracy,

TABLE 2. Number of samples in datasets used in model training and evaluation.

Cancer-type-specific dataset						
Cancer class	Split A		Split B		Split C	
	Train	Test	Train	Test	Train	Test
OC	24	4	24	4	19	9
EC	30	7	30	7	26	11
NSCLC	300	97	75	322	274	123
GBM	75	9	75	9	58	26
BM	30	6	30	6	25	11
SARC	30	8	30	8	26	12
Location-specific dataset						
Cancer class	Split A		Split B		Split C	
	Train	Test	Train	Test	Train	Test
OC_EC	55	10	55	10	45	20
NSCLC	150	247	55	342	274	123
GBM_ BM	100	20	55	65	83	37
SARC	32	6	32	6	26	12

F1-score, and top-2 balanced accuracy which considers if the right prediction is among the model's two top guesses. The Python code used for model training and R-script describing the step-by-step input image creation are provided in Supplemental Material section of the manuscript.

To overcome imbalanced classes in the dataset, we initially experimented with Generative Adversarial Networks for data augmentation, but it did not yield promising results. Hence, we focused on redistribution of the data. We decreased the number of samples from over-represented classes in the training set and added them to the test set. This data splitting was performed with a random sampler set on different random seeds to ensure impartiality. To assess generalized class performance, a Location-specific dataset was created, amalgamating cancers from close body locations. This involved grouping OC and EC samples as OC_EC class, and GBM and BM as GBM_BM. Cancer-type and location-specific datasets were further divided into the train and test sets in three ways: Split A contained the highest number of samples from all classes, leaving only a small portion for testing. Split B was introduced to create a more balanced training dataset, transferring more samples for testing purposes. Split C was introduced to check model training performance when 30% of data from each class was used for testing purposes only. This approach of redistributing samples. All models were trained and tested using a 5-fold Stratified Cross-Validation [34]. An overview of the train/test data split for multiclass classification is shown in Table.2.

Included XAI method – SHAP is based on cooperative game theory. Its goal is to identify the marginal contribution of each feature to a prediction. Here, SHAP values illustrate how individual gene expression levels, transformed into image pixels, influence predictions. Positive SHAP values denote gene expression levels aligning with specific cancer classifications, negative values indicate rejection and neutral values suggest no influence of the gene on the classification.

III. RESULTS

Upon establishing the most accurate DL model architecture during the initial tests as RN9-implat, the following

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 3. Performance characteristics of selected multiclass models (RN9-implat). Metrics are calculated for all 5 cross-validation folds for each experiment. Standard deviation (std) for folds is also included. Values are shown as percentiles.

Dataset	Balanced accuracy: mean (std)	F1: mean (std)	TOP-2 balanced accuracy: mean (std)
Cancer-type-specific dataset	66.51 (6.27)	65.77 (6.64)	89.76 (4.22)
Location-specific dataset	90.50 (2.78)	68.39 (3.13)	98.70 (4.64)

experiments were performed across both datasets and all data splits. Initially, we pursued multiclass classification to gauge the model's proficiency in discerning multiple cancer types. Subsequently, we delved into between-class classification, examining the model's ability to differentiate specific pairings, shedding light on its particular strengths and potential areas for refinement. Finally, we employed XAI (Explainable Artificial Intelligence) through the SHAP method, aiming to demystify the model's decision-making processes, corroborate its alignment with recognized biological markers, and enhance its credibility. This holistic approach was crucial for an exhaustive assessment of our model's efficacy and interpretability.

A. MULTICLASS CLASSIFICATION

The random chance of guessing the correct class for the Cancer-type and Location-specific datasets was equal to 16.67% and 25%, respectively. While the accuracy of all trained models was usually high (∼80%), owing to a large number of easily distinguishable NSCLC samples, the balanced accuracy dropped to 66.51% over 5 folds in the case of Cancer-type-specific dataset and to 90.50% over 5 folds for Location-specific dataset. Results are shown in Table.3. Detailed multiclass classification metrics for all tested runs are shown in the supplementary material ("Appendix.1 – models metrics").

Covering top-2 balanced accuracy for the model showed a significant increase in metrics, especially for the cancer-type-specific dataset, where the mean accuracies were lower. The inherent organization of the location-specific dataset, which groups some of the toughest-to-distinguish classes, could potentially explain this observation. This can also be observed in Figure 2, which shows the class-specific metrics and confusion matrix for the most accurate model on the Cancer-type-specific dataset.

B. BETWEEN-CLASS CANCER TYPE CLASSIFICATION

To grasp the nuanced distinctions and overlaps among cancer types, we conducted between-class classifications across both datasets. Key metrics, such as balanced accuracy and F1 scores from the location-specific dataset, are presented in Table 4. Detailed classification metrics are shown in the supplementary material ("Appendix.1 – models metrics").

A pivotal discovery was the model's unparalleled ability to differentiate AD samples from gynecological cancers (OC_EC), with most models achieving an impressive

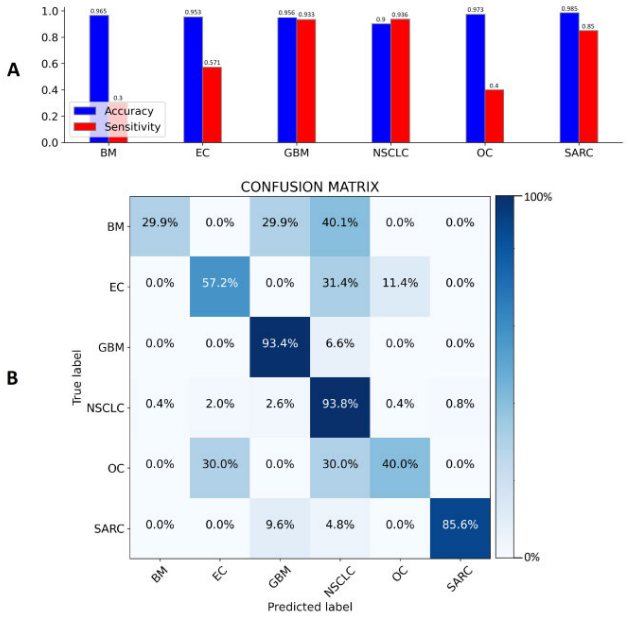

FIGURE 2. Performance results of the most-accurate RN9-implat model on test samples from the cancer-specific dataset. Values are calculated as a mean from all folds. (A) Accuracy and sensitivity were calculated for detecting every class apart from all samples. (B) Confusion matrix for test samples for each class. Numbers depict the percentage distribution of the classification of samples.

99% balanced accuracy across training, validation, and test sets. However, a decline in accuracy was observed when the models tackled brain-related cancers and sarcoma classifications. Further experiments with the cancer-type dataset enabled deeper class-to-class comparisons. Classification results are presented in Table 5. Detailed metrics are shown in the supplementary material ("Appendix.1 – models metrics").

As previously, our classifier showed outstanding performance distinguishing AD samples from OC and EC, reaching almost perfect scores on all tested data splits. Those two cancer types were also easily distinguished from all other cancer types except each other, which might be due to the fact that both cancer types are related to close locations.

C. EXPLAINABLE AI

Next, we applied the XAI method based on SHAP values on our trained models to get more insight into their performance. This experiment used samples of all distinct classes from the test split. The figure in the supplementary material ("Appendix.2 - XAI") shows the visualization of the outcome of applying SHAP on images fed to the multiclass model. Using the SHAP importance matrix for each pixel, we extracted 2-D coordinates of the top 5 highest positive values responsible for selecting the correct class and 2-D coordinates of the lowest 5 negative SHAP accountable for rejecting the incorrect class. Those coordinates were then back-translated into genes' names used to create input matrices. First, we tested the performance of binary classification

TABLE 4. Results of between-class classification of samples from the location-specific dataset. Numbers correspond to the mean balanced accuracy and F1 from cross-validation of models trained on data from classes in the row and column only.

Class	OC_EC		GBM_BM		NSCLC		SARC	
	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]
AD	99.89	99.50	88.96	77.31	96.36	95.34	88.49	91.75
OC_EC	-	-	99.60	99.40	87.81	88.79	98.79	98.34
GBM_BM	-	-	-	-	92.15	75.71	81.21	72.71
NSCLC	-	-	-	-	-	-	95.11	69.04

TABLE 5. Results of between-class classification of samples from the cancer-type-specific dataset. Numbers correspond to the mean balanced accuracy and F1 from cross-validation of models trained on data from classes in the row and column only.

Class	OC		EC		BM		GBM		NSCLC		SARC	
	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]	Balanced accuracy [%]	F1 [%]
AD	99.07	97.24	99.71	95.86	82.66	64.06	94.02	95.45	96.36	95.34	88.49	91.75
OC	-	-	60.00	57.72	96.67	96.64	99.47	98.74	81.31	79.76	95.00	93.02
EC	-	-	-	-	97.14	96.9	99.51	99.09	86.11	71.15	93.75	93.31
BM	-	-	-	-	-	-	71.47	72.62	90.86	63.59	90.42	90.84
GBM	-	-	-	-	-	-	-	-	94.75	91.93	88.38	86.07
NSCLC	-	-	-	-	-	-	-	-	-	-	95.11	69.04

TABLE 6. Genes with the most discriminating potential in classifying samples from cancer types and asymptomatic donors on the RN9-implat model. Arrows depict if the gene expression was higher or lower than the average for each gene from all samples.

Top genes for class:	Analyzed cancer type											
	OC		EC		BM		GBM		NSCLC		SARC	
AD	GNB1	↑	GSK3B	↑	FGFR4	↑	ITPR1	↑	MLLT4	↑	FZD2	↑
	NF2	↑	DDB2	↑	WNT7A	↑	ATF2	↑	GNB1	↑	RAC1	↓
	FGF22	↑	TAF4B	↑	MAPK13	↑	FGF4	↑	NF2	↑	HDDC3	↑
	MLLT4	↑	CPT2	↑	ADCYAP1R1	↑	PLA2G4C	↑	GNGT1	↑	GADD45A	↓
	NCBP1	↑	PIP4K2C	↑	FOS	↑	GABARAPL2	↓	CHUK	↑	SHMT1	↑
Cancer	AKR1C1	↑	SPTLC3	↑	POMT1	↑	LIG1	↑	HSPH1	↑	HSD3B1	↑
	CYP19A1	↑	SUFU	↑	B3GAT2	↑	FGF2	↑	IL17C	↑	UGT2A2	↑
	HBS1L	↑	MMP9	↑	COL1A1	↑	PDE2A	↑	ALDH9A1	↑	AMPD3	↑
	KRAS	↑	DUSP6	↑	ATF2	↑	PLA2G12B	↑	POMP	↑	SRP9	↓
	BTG3	↑	CYP3A4	↑	WNT11	↑	NDUFA9	↑	DUSP1	↑	MYLK	↓

models responsible for distinguishing between all the studied disease types and AD to discover which genes might be important when differentiating samples from healthy controls. Results are shown in Table 6.

Results show that a few genes (GNB1, NF2, MLLT4) were critical while displaying higher expression in healthy donors, helping to distinguish them from cancer patients. Genes correlating to detecting cancer were distinct for every class. Most of those genes are described in The Human Protein Atlas as prognostic markers or have strong evidence of being pathogenic. The subsequent phase involved evaluating multiclass models, aiming to pinpoint genes pivotal for accurate class identification. Results are shown in Table 7.

While there seem to be few exclusive genes in every class, in most instances, our model based its decision on combined information from the few most significant genes: DIAPH1, ACER2, BDKRB2, MLLT4, NF2, which were either upregulated or downregulated. Interestingly, MLLT4 was also substantial in AD samples while discriminating them from cancer patients.

TABLE 7. Genes that significantly impacted choosing the suitable class (bold) and rejecting false class in multiclass classification on the RN9-implat model. Arrows depict if the gene expression was higher or lower than the average for each gene from all samples.

Real class	Possible class											
	BM		EC		GBM		NSCLC		OC		SARC	
BM	**PCGF5**	↑	ACER2	↑	IKBKG	↑	DIAPH1	↑	ACER2	↑	FGF19	↑
	DIAPH1	↑	BDKRB2	↑	RAF1	↑	BDKRB2	↑	BDKRB2	↑	IRF7	↑
	KRAS	↓	DIAPH1	↑	SEC13	↑	ACER2	↑	PCGF1	↑	CCND1	↑
	ACER2	↑	MLLT4	↑	JUN	↑	MLLT4	↑	MAP2K2	↓	TNF	↑
	BDKRB2	↑	NF2	↑	CCND1	↑	TPCN1	↑	DIAPH1	↑	JAK3	↑
EC	PLCG2	↓	**ACER2**	↓	DIAPH1	↓	HSP90AA1	↓	HSP90AA1	↓	FGF19	↑
	MLLT4	↑	**LCN2**	↓	MLLT4	↑	IGF1	↑	IGF1	↑	PLCG2	↓
	BDKRB2	↑	**DIAPH1**	↓	PLCG2	↓	PHF5A	↑	PGK2	↑	MLLT4	↑
	TPCN1	↓	**PLCG2**	↓	BDKRB2	↑	IL4	↑	SHC3	↑	CCR10	↑
	DIAPH1	↓	**MLLT4**	↑	TPCN1	↓	GNG4	↑	TBK1	↓	BAX	↓
GBM	FGF23	↑	MLLT4	↑	**TPCN1**	↑	MLLT4	↑	SHC2	↑	PRKCG	↑
	MTMR14	↑	DIAPH1	↑	**NF2**	↑	DIAPH1	↑	MLLT4	↑	CCR10	↑
	PIK3R2	↑	SHC2	↑	**SHC2**	↑	TPCN1	↑	NF2	↑	PCGF1	↑
	PCGF5	↑	LCN2	↑	**MLLT4**	↑	HSPA5	↑	DIAPH1	↑	AP1B1	↓
	ENPP3	↑	NF2	↑	**DIAPH1**	↑	SHC2	↑	AIFM2	↑	CECR1	↑
NSCLC	CECR1	↑	FGF16	↑	NF2	↑	**MLLT4**	↑	FAM21C	↑	SHC2	↑
	NF2	↑	ITGA4	↑	SHC2	↑	**SHC2**	↑	PCGF1	↑	TBK1	↓
	SHC2	↑	LDHC	↑	DIAPH1	↓	**DIAPH1**	↓	BCL2	↑	NF2	↑
	TPCN1	↓	BCL2	↑	TPCN1	↓	**PHF5A**	↑	RAF1	↓	TPR	↓
	MLLT4	↑	EPS15	↓	MLLT4	↑	**NF2**	↑	PHF5A	↑	ENPP3	↑
OC	MLLT4	↑	ITGA4	↑	MLLT4	↑	IKBKE	↑	**BDKRB2**	↑	BAX	↓
	LCN2	↓	DDX46	↑	ACER2	↓	SNRPB	↓	**LCN2**	↓	MLLT4	↑
	ACER2	↓	S1PR5	↑	LCN2	↓	MTR	↑	**NF2**	↑	PLCG2	↓
	BDKRB2	↑	HSP90AA1	↑	NF2	↑	PPP2R3C	↓	**MLLT4**	↑	LCN2	↓
	NF2	↑	OPRD1	↑	BDKRB2	↑	FGF19	↑	**ACER2**	↓	FGF19	↑
SARC	KRAS	↑	BAX	↑	BAX	↑	BAX	↑	PIK3R3	↑	**CCR10**	↑
	JAK3	↑	LCN2	↑	FGF19	↑	FGF19	↑	HRAS	↑	**CASQ1**	↑
	TNF	↑	PGAM4	↑	IRF7	↑	IRF7	↑	BAX	↑	**FGF19**	↑
	PCGF5	↑	NF2	↑	JAK3	↑	BAD	↑	PLCG2	↑	**PIK3R3**	↑
	RELA	↑	MLLT4	↑	PAK1	↑	PLCG2	↑	PDGFD	↑	**BAX**	↑

IV. CONCLUSION

Developing a reliable, accessible, multiclass early cancer detection test is the holy grail of cancer diagnostics. Aiming to reach this goal, we compared common, convolution-based deep-learning models. Based on preliminary data, we selected ResNet architecture as the best solution to classify cancer using platelet RNA expression profiles. While the highest overall balanced accuracy reached by our multiclass models might still need to be improved for a standalone commercially available test, it could support the decision-making process of medical staff. We have observed that our model occasionally confuses brain metastatic cancer with glioblastoma and non-small cell lung cancer, and endometrial cancer with ovarian cancer. This is because these cancer types have similar locations or origins (many of the BMs originated from NSCLC) and may exhibit comparable RNA profiles in platelets. Narrowing the experiment down to 4 cancer-site-based classes showed a substantial increase in the model's performance, as the hardest-to-distinguish cancer types came from close locations. Our findings not only highlight the intricate interplay between platelets and cancer cells but also reveal a distinctive imprint of the tumor microenvironment and immune response within platelet RNA.

By reaching 66.51% balanced accuracy and 89.76% top-2 balanced accuracy for 6 types of cancer, we have established reliable competitors to other known research in terms of pan-cancer classifiers. Best et al. achieved comparable results with 71% accuracy and 89% top-2 accuracy when discriminating between healthy donors, glioblastoma, non-small cell lung cancer, colorectal cancer, pancreatic cancer, breast cancer, and hepatobiliary cancer [17]. However, samples from cancer location classes used in this research were collected from more diverse sites of origin. As we proved in this article - distinguishing between them is not as challenging. The latest pan-cancer study in platelets reported by [21] yielded 68% accuracy and 85% top-2 accuracy while classifying tumor sites of origin of 5 cancerous sites. Taking into account the number of classes used in research, our model proved more accurate and additionally allowed for further investigation with XAI methods.

Applying SHAP enriched our understanding of cancer's molecular intricacies, spotlighting the determinants influencing the classifier's decisions. A majority of these high-impact genes have been echoed in prior cancer studies, with many earmarked in The Human Protein Atlas as prognostic indicators for a plethora of cancers. The DIAPH1 has its role in

regulating the chromosomal instability of cancer cells as a growth inhibitor [35] and its overexpression was proven to be a predictive factor in breast cancer [36], and colorectal cancer [37]. The ACER2, as a key regulatory enzyme, promotes cancer cell survival [38] and can be used to predict molecular subtypes in bladder cancer [39]. The BDKRB2 role in cell angiogenesis was proven in cervical cancer [40], and breast cancer [41], and it was used as an epithelial-to-mesenchymal transition biomarker in glioma [42]. The MLLT4 was used among other genes in the prognosis of glioblastoma (downregulated) [43] and leukemia (upregulated) [44]. While XAI unveils potential avenues for classifier enhancement, it simultaneously validates their efficacy, reinforcing that their predictions are anchored in previously identified features, thus paving the way for personalized medicine.

One of this study's most significant limitations is the data availability. Whereas liquid biopsies are increasingly gaining popularity, Tumor Educated Platelets are an upcoming field with a limited amount of publicly available datasets. This scarcity, particularly for rarer cancer forms, is amplified in the context of deep learning models that voraciously consume data, potentially leading to overfitting as a smaller number of samples can result in an overly simplified representation and may lack generalization. Thus, we had to implement various sampling approaches and training on data split by a random seed to overcome this obstacle. As platelet RNA is gaining popularity in the liquid biopsy setting, it is likely that more datasets will be made publicly available in the future. It is also important to emphasize that each type of liquid biopsy material provides us with slightly different information on the patient's status. The use of a particular sample will be highly dependent on laboratory equipment used in a given unit and the expertise of its personnel. While circulating tumor cells (CTCs) might allow for in-depth analysis of multiple cancer cell clones, circulating tumor DNA (ctDNA) will be easier to obtain. However, while ctDNA, in certain circumstances, allows for more detailed profiling of the genetic properties of the tumor, its sensitivity may be lower as its abundance varies significantly, depending on the type of malignancy and the stage of the disease. Given the recent developments in the multiomics era, in the ideal scenario, liquid biopsies could combine different types of material (CTCs, ctDNA, TEPs, etc.). Unfortunately, most of the publicly available datasets are limited to only one type of material suitable for liquid biopsies.

Another obstacle rests in the fragmented nature of comparative analysis in the domain – many studies use different types of cancer classes, and we see the lack of a reliable reference dataset for machine learning classification algorithm development. Hence, our approach, beyond just multiclass classification, encompassed individual class performance analysis, setting the stage for future researchers to evaluate the performance of each class separately. In the evolving landscape of cancer diagnostics, our work carves a niche, blending state-of-the-art AI with crucial biological insights, and we believe it lays a solid foundation for further exploration and innovation in the domain.

SUPPLEMENTAL MATERIAL

Supplementary material includes: Detailed model metrics for all experiments (''Appendix.1 – model metrics'') and, visualization of the outcome of applying SHAP on images fed to the multiclass model (''Appendix.2 - XAI''). Data, Python code, R-script for input image generation, and multi-class classification model weights used in the project are available at:

https://gitlab.com/jopekmaksym/deep_learning-based_multiclass_approach_to_cancer_classification_on_liquid_biopsy_data.git. The authors are happy to send the binary classification model weights upon request.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of the following specialists from the Centre of Biostatistics and Bioinformatics at the Medical University of Gdańsk: Peter Grešner, Ph.D. (senior biostatistician) and Kamil Myszczyński, Ph.D. (senior bioinformatician). Thomas Würdinger is an inventor of relevant patents. He is a Shareholder of Illumina, Pacific Biosciences, and Oxford Nanopore. Maksym A. Jopek, Krzysztof Pastuszak, Sebastian Cygert, Myron G. Best, Jacek Jassem, Anna J. Żaczek, and Anna Supernat have no competing interests. They declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] H. Sung et al., ''Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,'' *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] P. Charkhchi, C. Cybulski, J. Gronwald, F. O. Wong, S. A. Narod, and M. R. Akbari, ''CA125 and ovarian cancer: A comprehensive review,'' *Cancers*, vol. 12, no. 12, p. 3730, Dec. 2020, doi: 10.3390/cancers12123730.

[3] O. Arrieta et al., ''Usefulness of serum carcinoembryonic antigen (CEA) in evaluating response to chemotherapy in patients with advanced non small-cell lung cancer: A prospective cohort study,'' *BMC Cancer*, vol. 13, no. 1, p. 254, Dec. 2013, doi: 10.1186/1471-2407-13-254.

[4] C. Alix-Panabières and K. Pantel, ''Liquid biopsy: From discovery to clinical application,'' *Cancer Discovery*, vol. 11, no. 4, pp. 858–873, Apr. 2021, doi: 10.1158/2159-8290.cd-20-1311.

[5] Q. Zhang, X. Song, and X. Song, ''Contents in tumor-educated platelets as the novel biosource for cancer diagnostics,'' *Frontiers Oncol.*, vol. 13, Apr. 2023, Art. no. 1165600, doi: 10.3389/fonc.2023.1165600.

[6] Y. N. Lamb and S. Dhillon, ''Epi proColon 2.0 CE: A blood-based screening test for colorectal cancer,'' *Mol. Diagnosis Therapy*, vol. 21, no. 2, pp. 225–232, Apr. 2017, doi: 10.1007/s40291-017-0259-y.

[7] A. M. Aravanis, M. Lee, and R. D. Klausner, ''Next-generation sequencing of circulating tumor DNA for early cancer detection,'' *Cell*, vol. 168, no. 4, pp. 571–574, Feb. 2017, doi: 10.1016/j.cell.2017.01.030.

[8] K. S. Thress et al., ''EGFR mutation detection in ctDNA from NSCLC patient plasma: A cross-platform comparison of leading technologies to support the clinical development of AZD9291,'' *Lung Cancer*, vol. 90, no. 3, pp. 509–515, Dec. 2015, doi: 10.1016/j.lungcan.2015.10.004.

[9] R. B. Lanman et al., ''Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA,'' *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0140712, doi: 10.1371/journal.pone.0140712.

[10] T. Cowling and H. Loshak, "An overview of liquid biopsy for screening and early detection of cancer," in *CADTH Issues in Emerging Health Technologies*. Ottawa, ON, USA: Canadian Agency for Drugs and Technologies in Health, 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK555478/

[11] K. A. Hoadley et al., "Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin," *Cell*, vol. 158, no. 4, pp. 929–944, Aug. 2014, doi: 10.1016/j.cell.2014.06.049.

[12] X. Zhang, J. Zhang, K. Sun, X. Yang, C. Dai, and Y. Guo, "Integrated multi-omics analysis using variational autoencoders: Application to pan-cancer classification," in *Proc. IEEE Int. Conf. Bioinformatics Biomedicine (BIBM)*, San Diego, CA, USA, Nov. 2019, pp. 765–769, doi: 10.1109/BIBM47256.2019.8983228.

[13] G. Di Sario et al., "Enhancing clinical potential of liquid biopsy through a multi-omic approach: A systematic review," *Frontiers Genet.*, vol. 14, Apr. 2023, Art. no. 1152470, doi: 10.3389/fgene.2023.1152470.

[14] A. Safrastyan and D. Wollny, "Network analysis of hepatocellular carcinoma liquid biopsies augmented by single-cell sequencing data," *Frontiers Genet.*, vol. 13, Aug. 2022, Art. no. 921195, doi: 10.3389/fgene.2022.921195.

[15] C. Alix-Panabières, "The future of liquid biopsy," *Nature*, vol. 579, p. 9, Mar. 2020, doi: 10.1038/d41586-020-00844-5.

[16] M. Lukasiewicz et al., "Diagnostic accuracy of liquid biopsy in endometrial cancer," *Cancers*, vol. 13, no. 22, p. 5731, Nov. 2021, doi: 10.3390/cancers13225731.

[17] M. G. Best et al., "RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics," *Cancer Cell*, vol. 28, no. 5, pp. 666–676, Nov. 2015, doi: 10.1016/j.ccell.2015.09.018.

[18] K. Pastuszak et al., "ImPlatelet classifier: Image-converted RNA biomarker profiles enable blood-based cancer diagnostics," *Mol. Oncol.*, vol. 15, no. 10, pp. 2688–2701, Oct. 2021, doi: 10.1002/1878-0261.13014.

[19] Y. Gao et al., "Platelet RNA enables accurate detection of ovarian cancer: An intercontinental, biomarker identification study," *Protein Cell*, Nov. 2022, Art. no. pwac056, doi: 10.1093/procel/pwac056.

[20] A. Supernat et al., "Transcriptomic landscape of blood platelets in healthy donors," *Sci. Rep.*, vol. 11, no. 1, p. 15679, Aug. 2021, doi: 10.1038/s41598-021-94003-z.

[21] G. J. G. Sjors et al., "Detection and localization of early- and late-stage cancers using platelet RNA," *Cancer Cell*, vol. 40, no. 9, pp. 999–1009, Sep. 2022, doi: 10.1016/j.ccell.2022.08.006.

[22] M. Kanehisa, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.

[23] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017, *arXiv:1705.07874*.

[24] M. G. Best, S. G. J. G. In 't Veld, N. Sol, and T. Wurdinger, "RNA sequencing and swarm intelligence–enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA," *Nature Protocols*, vol. 14, no. 4, pp. 1206–1234, Apr. 2019, doi: 10.1038/s41596-019-0139-5.

[25] N. Sol et al., "Blood platelet RNA enables the detection of multiple sclerosis," *Multiple Sclerosis J.-Experim., Translational Clin.*, vol. 6, no. 3, Jul. 2020, Art. no. 205521732094678, doi: 10.1177/2055217320946784.

[26] M. G. Best et al., "Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets," *Cancer Cell*, vol. 32, no. 2, pp. 238–252, Aug. 2017, doi: 10.1016/j.ccell.2017.07.004.

[27] K. M. Heinhuis et al., "RNA-sequencing of tumor-educated platelets, a novel biomarker for blood-based sarcoma diagnostics," *Cancers*, vol. 12, no. 6, p. 1372, May 2020, doi: 10.3390/cancers12061372.

[28] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol.*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.

[29] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron, "Variance stabilization applied to microarray data calibration and to the quantification of differential expression," *Bioinformatics*, vol. 18, pp. 96–104, Jul. 2002, doi: 10.1093/bioinformatics/18.suppl_1.s96.

[30] W. Luo, M. S. Friedman, K. Shedden, K. D. Hankenson, and P. J. Woolf, "GAGE: Generally applicable gene set enrichment for pathway analysis," *BMC Bioinf.*, vol. 10, no. 1, p. 161, May 2009, doi: 10.1186/1471-2105-10-161.

[31] S. Cygert et al., "Platelet-based liquid biopsies through the lens of machine learning," *Cancers*, vol. 15, no. 8, p. 2336, Apr. 2023, doi: 10.3390/cancers15082336.

[32] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[34] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 542–545, doi: 10.1016/B978-0-12-809633-8.20349-X.

[35] S. Miao, P. Schäfer, J. Nojszewski, F. Meyer, and S. Windhorst, "DIAPH1 regulates chromosomal instability of cancer cells by controlling microtubule dynamics," *Eur. J. Cell Biol.*, vol. 100, no. 3, Apr. 2021, Art. no. 151156, doi: 10.1016/j.ejcb.2021.151156.

[36] J. Mei et al., "Formin protein DIAPH1 positively regulates PD-L1 expression and predicts the therapeutic response to anti-PD-1/PD-L1 immunotherapy," *Clin. Immunol.*, vol. 246, Jan. 2023, Art. no. 109204, doi: 10.1016/j.clim.2022.109204.

[37] Y. Lin et al., "Expression of DIAPH1 is up-regulated in colorectal cancer and its down-regulation strongly reduces the metastatic capacity of colon carcinoma cells," *Int. J. Cancer*, vol. 134, no. 7, pp. 1571–1582, Apr. 2014, doi: 10.1002/ijc.28486.

[38] B. Liu, J. Xiao, M. Dong, Z. Qiu, and J. Jin, "Human alkaline ceramidase 2 promotes the growth, invasion, and migration of hepatocellular carcinoma cells via sphingomyelin phosphodiesterase acid-like 3B," *Cancer Sci.*, vol. 111, no. 7, pp. 2259–2274, Jul. 2020, doi: 10.1111/cas.14453.

[39] J. Liu et al., "ACER2 forms a cold tumor microenvironment and predicts the molecular subtype in bladder cancer: Results from real-world cohorts," *Frontiers Genet.*, vol. 14, Mar. 2023, Art. no. 1148437, doi: 10.3389/fgene.2023.1148437.

[40] Y. Zhou et al., "Serum bradykinin levels as a diagnostic marker in cervical cancer with a potential mechanism to promote VEGF expression via BDKRB2," *Int. J. Oncol.*, vol. 55, no. 1, pp. 131–141, May 2019, doi: 10.3892/ijo.2019.4792.

[41] C. Dubuc et al., "Targeting intracellular b2 receptors using novel cell-penetrating antagonists to arrest growth and induce apoptosis in human triple-negative breast cancer," *Oncotarget*, vol. 9, no. 11, pp. 9885–9906, Feb. 2018, doi: 10.18632/oncotarget.24009.

[42] Y. Yang, J. Wang, F. Shi, A. Shan, S. Xu, and W. Lv, "BDKRB2 is a novel EMT-related biomarker and predicts poor survival in glioma," *Aging*, vol. 13, no. 5, pp. 7499–7516, Mar. 2021, doi: 10.18632/aging.202614.

[43] H. Yang, L. Jin, and X. Sun, "A thirteen-gene set efficiently predicts the prognosis of glioblastoma," *Mol. Med. Rep.*, vol. 19, no. 3, pp. 1613–1621, Jan. 2019, doi: 10.3892/mmr.2019.9801.

[44] X. Chen et al., "Retrospective analysis of 36 fusion genes in 2479 Chinese patients of de novo acute lymphoblastic leukemia," *Leukemia Res.*, vol. 72, pp. 99–104, Sep. 2018, doi: 10.1016/j.leukres.2018.08.009.