_____

Bioinformatics and Machine Learning in Skin Cancer Risk Assessment and Prognosis: A Review

D. Surendren[1], J. Sumitha[2]

[1] Research Scholar, Department of Computer Science,
Dr. SNS Rajalakshmi College of art and science, Coimbatore, Tamil Nadu, India
[1] surendrenmca@gmail.com

[2] Assistant Professor, Department of Computer Science,
Dr. SNS Rajalakshmi College of art and science, Coimbatore, Tamil Nadu, India
[2] sumivenkat2006@gmail.com

Abstract—Skin cancer is one of the leading dangerous varieties of cancer. Carcinoma Skin cancer is caused by unrepaired deoxyribonucleic acid (DNA) in skin cells that produce mutations or genetic defects on the skin. Carcinoma has a tendency to moderately extend to other parts of body, therefore it is easy to cure during early stages, and hence it is required to detect as early as possible. Every year, doctors diagnose carcinoma in around three million or more patients across the world. Nowadays, it is one of the most widely recognized forms of cancers for human health. Hence, we need an early diagnosis to prevail any crucial condition of the infected patients. There a lot of factors such as the rate of increase of cases, increased death rate, and more expensive and painful medical treatment. Having considered the seriousness of those problems, researchers have developed numerous early finding methods for skin cancer. Despite clinical staging guidelines, the prognosis skin cancer (metastatic melanoma) is variable and difficult to predict. Machine Learning and Bioinformatics take inputs from clinical, histopathology and genetic to analyze to predict risk with high accuracy of melanoma patients. This literature review aims to provide key genetics science drivers of malignant melanoma and up to date applications of machine learning models and bioinformatics with the risk discovery of carcinoma patients. A robustly valid risk stratification tool will probably guide the medical practitioner management of malignant melanoma patients and ultimately improve patient outcomes. Review findings are presented in tables for better understanding.

Keywords- Bioinformatics, machine learning, melanoma, skin cancer.

I. INTRODUCTION

Cancer is a cell based disease and it affects the cells. The basic building block of body is cell. The other organs in the body and tissues are made of cells. The body produces new cells continuously, which makes us to grow, also replaces the dead cells and heal any injures. Generally, the cells die in an orderly fashion, so that those cells shall be replaced with new cells. Sometimes, there may be a abnormal cell growth and that abnormal growth may lead to cancer. The condition of abnormal cell growth within skin is referred as Carcinoma. The skin cancer can be categorized into three major types. They are melanoma, SCC (Squamous Cell Carcinoma), BCC (Basal Cell Carcinoma. The later types SCC and BCC are also known as non-melanoma carcinoma or keratinocyte cancer. Non-melanoma is more regular than melanoma [1].

In the United States, information about BCC and SCC is not actively gathered in population based central cancer repositories. Instead the details about skin cancer are got through studies, surveys and from medical claim reports. On an average about 5 million skin cancer patients are treated every year. As per Agency for Healthcare Research and Quality's Medical Expenditure Panel Survey, there are about 4.3 million adults are treated for different types of skin cancers. Melanoma is considered to the most familiar type of skin cancer which causes most of the casual death. As per the federal data for 2007–2011, around 63,000 U.S citizens are detected with melanoma, and roughly 9,000 patients decease on a year. According to the available data, mostly melanoma spread was most common among the elder adults and very less proportionate among youngsters and children [2].

Computer Algorithms have many applications; Machine Learning (ML) is one such application with the objective to improve the prediction accuracy of the concerned problem [3]. ML algorithms are based on pattern recognition and are designed improve its behaviour based on data or experience, without additional human intervention. These algorithms can be powerful tools to aid humans in the analysis of large, heterogeneous data sets, such as genomic data sets. The bioinformatics analysis of genomic data has been extensively applied to identify potential genetics and signaling pathways associated with melanoma pathogenesis and metastasis. More recently, bioinformatics analyses, including machine learning, are increasingly utilized to predict prognosis, risk stratify, and ultimately inform personalized treatment in cutaneous melanoma [14].

We have done a literature survey within Google Scholar, Science Direct, PubMed and many Reputed Journals to give a general idea about the applications of machine learning and bioinformatics in skin cancer prediction and risk stratification. In this survey we tried to recapitulate the recently setup key factors of carcinoma that utilize bioinformatics for finding. We also summarize the algorithms, key findings, and prognosticative precision of current studies using machine learning algorithms and bioinformatics to carcinoma risk stratification.

Section 2 is an overview of the Algorithms for Machine Learning in Skin Cancer; Section 3 is Machine Learning and Bioinformatics in Cancer Risk Assessment and provides the summary of machine learning and bioinformatics in skin

_____

cancer prognosis. The last Section explains the conclusion part of this research.

II. MACHINE LEARNING ALGORITHM IN SKIN CANCER

Because of the elevated dominance of skin cancer rate, an advanced range of individuals need prompt prognosis, diagnosing and on-going observation. This put more pressure on expert medical examination, which can be allayed by higher patient self-surveillance methods and also the use of expert systems for medical practitioners who have very less experience. The diagnosis done by machines is basically objective, and is not influenced by external factors. On the other hand, the diagnosis done by human is associated with subjective variations and may be influenced by some of the external factors. Machine learning algorithms are based on pattern recognition and are designed improves its behavior based on data or experience, without additional human intervention. These algorithms can be powerful tools to aid humans in the investigation of large, heterogeneous data sets, such as genomic data sets. The number of biopsies considered can be reduced significantly if the early detection system is implemented with the help of Artificial Intelligence techniques and with correct regulatory measures. If implemented with the necessary regulations, the use of Artificial Intelligence for the detection and progression of carcinoma cancer may result in fewer biopsies. By providing a proper assistance, skin cancer patients and their attenders can make self-examination of skin with less intervention from the experts

III. BIO INFORMATICS AND MACHINE LEARNING IN CANCER RISK ASSESSMENT

In spite of clinical staging guidelines, predicting the prognosis of melanoma is challenging due to its heterogeneous nature. Bioinformatics tools have been predominantly used to analyze NGS data and help identify potential mutations connected with malignancies pathogenesis [16]. In recent days, there has been tremendous increase in applications of bioinformatics analysis in stratification of melanoma risk and the prognosis prediction to give information about treatment. The melanoma of Stage III and stage IV require approved appropriate therapies which are currently used extensively after the resection of highly developed melanoma. [17]. On the other hand, these general therapies are connected with undesirable events of frequent grade 3 or 4, and are expensive. As such, perfect explorative tools to forecast the probability of reappearance and endurance are required to stratify risk to better recognize suitable candidates for additional treatment and level of supervision. The summary of the study is given below table 1

Table 1. Summary of bioinformatics and machine learning in skin cancer prognosis

#	Author	Methods	Key Findings	Performance
1	Anjali Dhall, et al 2020	Decision Tree Lasso	A Regression Model was developed for predicting Overall	-
	[12]	Random Forest, Ridge	Survival Rate in Cutaneous Melanoma affected Patients.	
2	Yabin Cheng, et al. 2015 [4]	Cox proportional multivariate analyses and Kaplan-Meier survival	BRAF and MMP2 are the best predictive biomarkers for AJCC Stages I and II (P = 0.010, 0.000, Log-rank test), respectively, but p27 is a good marker for AJCC Stages III/IV (P = 0.010, 0.000, Log-rank test) (0.018, 0.046, respectively, log-rank test)	Proportions for Stages I, II, III, and IV are 89.0%, 61.0%, 40.6% and 8.2%
2	ChakitArora, et al. 2020 [13]	SVR Based Regression Model	Developed a web server and an Android application called "CMcrpred" to predict a melanoma patient's survival outcome and risk.	SVR Based Regression Model HR 3.19
4	David Rofman, et al. 2018 [6]	Artificial Neural Networks	Developed a Multi-Parameterized Artificial Neural Network: Predict and Stratify an Individual's Risk of Developing Non-Melanoma Skin Cancer (NMSC) based on Health Informatics.	BCC (Sensitivity: 40%, Specificity :87%), SCC (Sensitivity: 39%, Specificity : 92%)
5	Denis Shepelin, et al. 2015 [5]	Support Vector Machine, Random Forest, Boosted, Naive Bayes Classifier, Logistical Regression and Partial Least Squares	At the level of molecular pathway activation, developed a stable model describing the creation and progression of melanoma. They discovered 44 signalling and metabolic pathways that were linked to the production of nevi, the progression of primary melanoma, and its metastases.	Accuracy >= 89%.
6	GoranJonsson, el al.	Support Tree Analys	Four distinct subtypes with distinct gene signatures are linked	-

_____

No	Author	Technique	Description	Result
	2010 [19]	is , Two-group Significance of Microarray Analysis, Unsupervised hierarchical clustering,	to various prognoses.	
7	Hsiao-Han Wang, et al. 2019 [7]	Convolutional Neural Networks	A Risk based Prediction Model was developed to Predict the factors for non-melanoma Skin Cancer.	AUROC : 0.894, with a sensitivity of 83.1% and specificity of 82.3%
8	Manik Garg, et al. 2021 [10]	Classification Models using Machine Learning (Random Forest Classifier)	ML models trained using metastasis-associated genes outperformed models trained with clinical variables or published prognostic markers in predicting regional lymph node metastasis.	pAUROC < 0.05
9	M. Lingaraj, et al. 2021 [20]	Veritable Support Vector Machine	A Novel Algorithm was designed to predict the existence of Melanoma	HIS2828 : 82.11%, ISIC2017: 88.10%
10	Paul Johann et, et al 2018 [11]	Multivariable Classifier (Kaplan–Meier)	The Resulting Classifier Accurately Layered Patients into High versus Low Risk for Disease Progression.	AUC: 0.805
11	Pooya Mobadersany, et al. 2018 [21]	Survival Convolutional Neural Networks,	Developed A Genomic Survival Convolutional Neural Network. By incorporating genomic data into the fully connected layers of the Survival Convolutional Neural Network, it learns from both genomics and histology at the same time.	-
12	Sherry Bhalla, et al. 2019 [9]	Support Vector Classification	Developed a Machine Learning Models to Distinguish mRNA, miRNA and methylation data of TCGA.	AUROC: 0.95, Accuracy: 89.47%
13	Xijia Chen, et al. 2016 [8]	Kaplan-Meier Method	The Prognostic Signature of four-lncRNA was identified which shall be used in risk predication of melanoma affected patients.	Accuracy >= 84%.
14	Zufeng Sheng, et al. 2020 [18]	Kaplan-Meier curves	The Genes which are prognostic differently expressed are identified among normal and metastatic melanoma.	-

IV. CONCLUSION

Here we present the outline of the end results of the literature survey. Skin cancer is a heterogeneous condition with many patients subgroups linked with different possible outcomes. Bioinformatics techniques analyses, significantly machine learning is internally assessed to accurately identify skin cancer patients at risk. Nevertheless, for having clinical efficacy, bioinformatics tools should be rigorously. Bioinformatics techniques and ML algorithm analyses are increasing quickly in the field of skin cancer prognosis, and we expect that unremitting research in skin cancer risk screening tools can likely to change the prospect patient management system and outcomes. In many research finding process Kaplan-Meier, Support Vector Machine and Random Forest Classifier algorithms are used for predicting and diagnosing skin cancer. From the literature summary the support vector machine classification algorithm provide more accuracy than the other machine learning algorithms. It gives more than 88% accuracy in many cases.

REFERENCES

1. https://www.cancer.org.au/assets/pdf/understanding-skin-cancer-booklet (2022).
2. https://www.hhs.gov/sites/default/files/call-to-action-prevent-

_____

skin-cancer.pdf (2022).

3. Maxwell W. Libbrecht1 and William Stafford Noble, "Machine learning applications in genetics and genomics", NATURE REVIEWS, GENETICS, Vol 16, pp. 321-332, (2015).

4. Yabin Cheng, Jing Lu, Guangdi Chen, "Gholamreza Safaee Ardekani, et.al, Stage-specific prognostic biomarkers in melanoma", Oncotarget, Vol. 6, No.6, pp. 4180-4189, (2015).

5. Denis Shepelin, Mikhail Korzinkin, Anna Vanyushina, Alexander Aliper, et al. "Molecular pathway activation features linked with transition from normal skin to primary and metastatic melanomas in human", Oncotarget, Vol. 7, No. 1, pp.656-670, (2015).

6. David Rofman, Gregory Hart, Michael Girardi, Christine J. Ko & Jun Deng, "Predicting Non-Melanoma Skin Cancer Via A Multi-Parameterized Artifcial Neural Network", Scientific Reports (2018).

7. Hsiao-Han Wang, MD; Yu-Hsiang Wang; Chia-Wei Liang, MS; Yu-Chuan L, "Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer", JAMA Dermatol, 9 American Medical Association (2019).

8. Xijia Chena , Wenna Guoa,c , Xin-jian Xub , Fangchu Su et al, "Melanoma long non-coding RNA signature predicts prognostic survival and directs clinical risk-specific treatments", Journal of Dermatological Science (2016).

9. Sherry Bhalla1,Harpreet Kaur, Anjali Dhall & Gajendra P. S. Raghava, "Prediction and Analysis of Skin Cancer Progression using Genomics Profles of Patients", Scientific Reports (2019).

10. Manik Garg, Dominique-Laurent Couturier, Jeremie Nsegimana, Nuno A Fonseca, et al, "Tumour gene expression signature in primary melanoma predicts long-term outcomes", Nat Commun, Vol. 12(1) (2021).

11. Paul Johannet, Nicolas Coudray, Douglas M. Donnelly, George Jour, et al, "Using Machine Learning Algorithms to Predict Immunotherapy Response in Patients with Advanced Melanoma", CLINICAL CANCER RESEARCH, American Association for Cancer Research (2020).

12. Anjali Dhall, Sumeet Patiyal, Harpreet Kaur, Sherry Bhalla, et al, "Computing Skin Cutaneous Melanoma Outcome From the HLA-Alleles and Clinical Characteristics", Frontiers in Genetics (2020).

13. Chakit Arora, Dilraj Kaur, Anjali Lathwal, Gajendra P.S. Raghava, "Risk Prediction in Cutaneous Melanoma Patients fromtTheir Clinico-Pathological Features: Superiority of Clinical Data over Gene Expression Data", Heliyon (2020).

14. Emily Z. Ma, Karl M. Hoegler and Albert E. Zhou, "Bioinformatic and Machine Learning Applications in Melanoma Risk Assessment and Prognosis: A Literature Review", Genes (2021).

15. J. Sumitha,T. Devi &D. Ravi, "Comparative Study on Gene Expression for Detecting Diseases Using Optimized Algorithm", International Journal of Human Genetics, Voi(17), pp.38-42, (2017).

16. Olga Papadodima, Georgia Kontogianni, Georgia Piroti, Ilias Maglogiannis, et.al, Review, "Genomics of cutaneous melanoma: focus on next-generation sequencing approaches and bioinformatics", Journal of Translational Genetics and Genomics (2018).

17. Reinhard Dummer, Axel Hauschild, Mario Santinami, Victoria Atkinson, et. al, "Five-Year Analysis of Adjuvant Dabrafenib plus Trametinib in Stage III Melanoma", The New England Journal of Medicine (2020).

18. Zufeng Sheng, Wei Han, Biao Huang, Guoliang Shen, "Screening and identification of potential prognostic biomarkers in metastatic skin cutaneous melanoma by bioinformatics analysis", Journal of Cellular Molecular Medicine (2020).

19. Goran Jonsson, Christian Busch, Stian Knappskog, Jurgen Geisler, et. al, "Gene Expression Profiling–Based Identification of Molecular Subtypes in Stage IV Melanomas with Different Clinical Outcome", Human Cancer Biology, Clinical Cancer Research (2010).

20. M.Lingaraj, A.Senthilkumar, J.Ramkumar, "Prediction of Melanoma Skin Cancer Using Veritable Support Vector Machine", Annals of R.S.C.B., Vol. 25, Issue 4, Pp. 2623–2636 (2021).

21. Napoleon D. and Praneesh M. "Detection of Brain Tumor using Kernel Induced Possiblistic C-Means Clustering", volume no.3, issue no.9, pp 436-438, (2013).

: