THE STUDY OF TISSUE HETEROGENEITY AND CLASSIFICATION USING AI TECHNIQUES

An Undergraduate Research Scholars Thesis

by

JUDE ALOUDEH[1] AND MOHAMED ZEID[2]

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Othmane Bouhali

May 2021

Majors: Electrical Engineering[1&2]

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

We, Jude AlOudeh and Mohamed Zeid, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Faculty Advisor prior to the collection of any data used in this final thesis submission.

This project required approval from the Texas A&M University Research Compliance & Biosafety office.

IRB exempted from HMC since it is retrospective. MRC identification: MRC-01-19-002.

TABLE OF CONTENTS

Page

ABSTRACT

The Study of Tissue Heterogeneity and Classification Using AI Techniques

Jude AlOudeh and Mohamed Zeid
Department of Electrical and Computer Engineering
Texas A&M University

Research Faculty Advisor: Dr. Othmane Bouhali
Department of Electrical and Computer Engineering
Texas A&M University

The idea behind our project is to design an algorithm that utilizes artificial intelligence to detect tissue heterogeneity in patients without the need to carry out an invasive biopsy. We aim to make the cancer prognosis process based solely on the study of the scanned medical images such as MRI or CT. The algorithm will be written in Python and will utilize large data sets of radiomics biomarkers extracted from medical images of different modalities through a software called LIFEx. Radiomics biomarkers are huge amounts of quantitative features extracted from medical images that characterize tumor phenotypes like texture and shape. The objective that we want our algorithm to achieve is to classify the cancer stage. In this project, we will focus on cervix cancer as it is of great interest to our collaborators who are providing us with private data. Another benefit to our algorithm is that it will offer a noninvasive method for cancer diagnosis and will hence bypass biopsies as they are associated with many additional health risks and costs. This project will contribute to changing the way doctors diagnose cancer and make it a more efficient process using our robust, reliable detection of tissue heterogeneity.

ACKNOWLEDGEMENTS

Contributors

We would like to thank our faculty advisor, Dr. Othmane Bouhali, and our medical collaborators, Dr. Tarraf Torfeh and Dr. Souha Aouadi, for their guidance and support throughout the course of this research.

Thanks also go to our friends and colleagues and the department faculty and staff for making our time at Texas A&M University a great experience.

Finally, thanks to our other team member for their help and contribution in this research.

The cervical cancer patient data used for model training was provided by Dr. Tarraf Torfeh and Dr. Souha Aouadi.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

This Undergraduate research did not receive any funding.

1. INTRODUCTION

Cervical cancer has been determined to be the fourth leading cancer induced cause of death in developed countries, and the second most common cause of death (due to cancer) in developing countries. It is also the second most occurring form of cancer among women. Moreover, there are 530,000 new cervical cancer cases worldwide every year [1].

The treatment method for patients suffering from cervical cancer differs depending on the tumor stage and nodal analysis. For early stages of cervical cancer with locally confined tumors, the treatment strategy involves surgically removing a part or all of a tissue, structure, or organ where the cancer exists. For more advanced stages of cancer with node-positive tumors, more extensive methods are required, including radio chemotherapy [2]. Federation of Gynecology and Obstetrics (FIGO), is a cervical cancer staging system introduced in 2018 [3]

It consists of five different stages [4]:

- stage 0: carcinoma in situ (common in cervical, vaginal, and vulval cancer)
- stage I: confined to the organ of origin
- stage II: invasion of surrounding organs or tissue
- stage III: spread to distant nodes or tissue within the pelvis
- stage IV: distant metastasis(es)

Several studies have shown that cervical cancer has a five-year survival rate of 65% for FIGO stage II, 40% for FIGO stage III , and 15% for FIGO stage IV-A [2].

A method that has been proven to improve the result of the treatments of cervical cancer is early diagnosis. There are a few biomarkers that relate to cervical cancer such as FIGO stage, histology, tumor volume, lymph node metastasis (LNM), and single gene markers. These are all

used as prognostic factors when characterizing the cancer [2] . With the use of these biomarkers, decisions on treatment strategies can be made. In addition to the biomarkers discussed above, feature extraction from various imaging modalities can be used to assist in decisions regarding treatment plans for patients. This is done using radiomics [5]; Radiomics transforms medical images into high-dimensional quantitative features using various data characterization algorithms [6]. Some of the most essential features such as tumor volume and major axis length can be extracted from MRI images. [7]

1.1 Diffusion Weighted Imaging

Diffusion-weighted MR imaging (DWI) measures the diffusion of water molecules within cellular tissues [8]. The contrast of DWI imaging is a result of the variance in the water molecules mobility in the different regions [10]. This makes DWI sensitive to smaller abnormalities in tissue, hence it is able to provide a more detailed characterization of the tissue [8]. Apparent diffusion coefficient (ADC) values are calculated from DWI images to quantitatively assess the magnitude of the diffusion of water molecules in the tissue [11].

The choice of b-values is crucial for the acquisition of DWI. This is decided based on the context in which the image is obtained as well as the model chosen for analysis. For the models based on both perfusion and diffusion information, a range of low b-values and high b-values are needed. For a voxel-by-voxel analysis, given very high b-values, the signals could be impacted by a high level of noise. In order to select the optimal b-values for a specific model and a specific anatomical site, a careful assessment of the signal decay curves is necessary [10].

In [8], a study is conducted to investigate the use and effectiveness of DWI in detecting and staging of cervical cancer before and after therapy. It still remains a challenge to radiologists to accurately assess the tumor’s response to therapy. DWI imaging can effectively display the

tumor due to its superior contrast with normal cervix tissue. Several previous studies were discussed in [8]; one study found that the mean ADC value of cervical cancer was 1.09±0.20×10−3mm2/s which turned out to be lower than that of normal cervix 1.79 ± 0.24 × 10−3 mm2 /s and increased after therapy to 1.48±0.23×10−3 mm2/s. Another study showed that the median ADC value of cervical cancer was 1.09 ± 0.20 × 10−3 mm2 /s and that is also lower than the ADC value of normal cervix identified by this study to be 2.09 ± 0.46 × 10−3 mm2 /s. It can be shown from the results of both studies that the value of ADC for cervix cancer is lower than that of normal cervix tissue. This reduction is expected to be due to hypercellularity within malignant tissues. These studies also determined that the completion of therapy results in increased ADC values [8].

1.2 Radiomics workflow

The first step in the radiomics workflow for determining the cancer stage of a patient is the image outcome of an MRI scan. Using this image outcome to calculate a corresponding Apparent Diffusion Coefficient (ADC) image is the first of the raw DW MRI data processing stages which offers a better diffusion representation. The raw data is then segmented in order to identify the Region Of Interest (ROI), either through the use of an automated segmentation program or by an experienced radiologists that manually delineates it. Using software such as LIFEx [12], the qualitative features that are present in the medical image are extracted from the ROI. Depending on the imaging modality of choice (or should this just be about DWI MRI?), said features could include the cellular composition characteristics of the ROI as well as its interaction with the surroundings, textural features as well as the shape of the tumor which is later used to classify the cancer stage. The following and final stage of the radiomics workflow is the machine learning model. The machine learning model starts by the data going through

feature reduction algorithms such as the Support Vector Machine with Recursive Feature Elimination (SVM-REF) or the Principal Component Analysis (PCA). These reduction algorithms are used to remove any and all heavily correlated features to improve the future accuracy of the classifier, leaving only the features most significant to the class that have the smallest p values ($p \leq 0.05$) [13]. The remaining, most significant features are then fed into a machine learning classifier that is able to determine the probabilities of the tumor outlined in the ROI being in the different stages. This classifying procedure is only possible after the machine learning algorithm of choice, be it a Neural Network (NN), Random Forest (RF), SVM, or any other classifier, is trained beforehand on an appropriate data set with the resulting accuracy tested on a testing set to ensure the algorithms accuracy.

1.3 Prediction of histological type

Winfield et. al conducted a study on the ability of different DW MRI models to differentiate between types and grades of cervical cancer tumors. The non-mono-exponential model deemed to be the best model to differentiate between tumor types based on its parameters: α (stretched exponential), K (kurtosis), f and D* (bi-exponential) which showed significant differences in tumor types. ADC alone is sufficient to predict the grade of the tumor, however, the use of the non-mono-exponential model parameters along with ADC can provide a better characterization of the tumor type, i.e. squamous cell carcinoma versus adenocarcinoma. Therefore, this study concludes that the non-mono-exponential model parameters describe different aspects of the tumor microstructure with the stretched exponential model showing uncorrelated parameters that describe the histological features of the tumor. This is therefore recommended to be the most relevant for prognosis prediction [14].

1.4 Prediction of tumor's grade

Tumor grade is important information that MDs use to decide on the tumor treatment. Cancer grading describes the cellular phenotype of the tumor in comparison to healthy cells, the treatment protocol differs from one grade to another which is why grade identification is a crucial step in cancer treatment decisions [15,16].

In a research conducted by Dr. Yin Liu which compared the grade evaluation accuracy between using Regions Of Interest (ROI) and Volumes Of Interest (VOI) for different b-values, it was concluded that VOI yielded better results and found an optimum b value of 1000 when dealing with DWI pelvic scans. The Least Absolute Shrinkage and Selection Operation (LASSO) regression was the regression technique of choice, 10-fold cross validation and overall misclassification were used to evaluate the performance of the different parameters as well as find the optimum λ (tuning/ penalty coefficient) for the LASSO. The features included histogram-based, textural, and Laplace of Gaussian filtered features. The results showed a slightly better performance for the b value of 1000 when compared to the b value of 800 for the VOI with a superior misclassification percentage of (0.3642 ± 0.0162) compared to (0.3758 ± 0.0118). While this paper demonstrated the superiority of VOI over ROI, its range of tested b values were very limited and hence did not demonstrate as strong a proof for the b value selection [9].

A paper published by Xi Zhang looked to distinguish between high and low grade bladder cancer using textural features extracted from DWI and ADC and proposed a radiomics based cancer grading strategy using textural features only. With a b value of a 1000, textural features were derived from Haralick and histogram features obtained from the VOIs. This paper demonstrated the superiority of Support Vector Machine with Recursive Feature Elimination (SVM-REF) over standard SVM classification [9].

1.5 Prediction of tumor's stage

The staging system used to classify the stage of cervical cancer is International Federation of Gynecology and Obstetrics (FIGO) (i.e., stages I, II, III, and IV). To characterize and correlate the FIGO stages of cervical cancers, whole-lesion ADC first-order statistics and texture features can be used after the MRI analysis. This allows for the comparison between lower FIGO (IB-IIA) cervical cancer stages and higher FIGO (IB-IIA) stages. [17] Some of the first order statistics that proved to be a very efficient method to differentiate between lower and higher stages of cervical cancer include skewness, kurtosis and entropy. [18] First order statistics are an ADC based feature and provide the distribution of ADC values throughout the image [17] [18]. The use of the first order statistics and texture features allows for Spearman's rank correlation coefficients to be calculated for the set of FIGO stages. The ManneWhitney U-test can be utilized to compare stages IB-IIA to IIB-VA cervical cancers. While the diagnostic performance of all whole-lesion ADC features can differentiate between IB-IIA from stage IIB-IVA cervical cancers. The diagnostic performance is done using a Receiver Operating Characteristic (ROC) curve analysis [18]. In ADC, first-order statistics comparing the skewness allows to differentiate between various FIGO stages, in specific stage IB or IIA from stage IIB, IIIA, or IIIB cervical cancers; the values for the skewness were skewness 1.4 and 2.3, respectively. Higher cervical cancer stages resulted in more positively skewed histograms. Kurtosis describes the tailedness of ADC histograms. Higher stages of cervical cancer produce a much larger kurtosis than that of lower stages meaning the distribution has more outliers. Entropy expresses the randomness of ADC values within a VOI. It was found to be that higher stages of cervical cancer have a much higher entropy than that of lower stages of cancer [18].

2. METHODS

2.1 Algorithm Workflow

The project is sectioned out into the following parts. The first step is data acquisition from various sources to train the algorithm. The second step is radiomics feature extraction which is done by the LifeX software. The third step is data splitting into multiple sets for the implementation of cross validation at a later stage. The fourth step constitutes feature reduction and sub-sampling. The fourth step is processing of the data. This is broken down into 4 main parts the first is the feature reduction and selecting data which is important to normalize the features and remove the heavily correlated ones. The second part is sub-sampling, this is done to ensure the optimum ratio between the different classes. The third is the classification stage which uses supervised learning. Finally, we have the evaluation stage, where we use the Area Under the Receiver Operating Characteristic Curve to compute the accuracy. The final step is classifying the grade of the cancer. This project focusses on cervical cancer, yet the same logic can be used in this algorithm can be implemented to different forms of cancer.

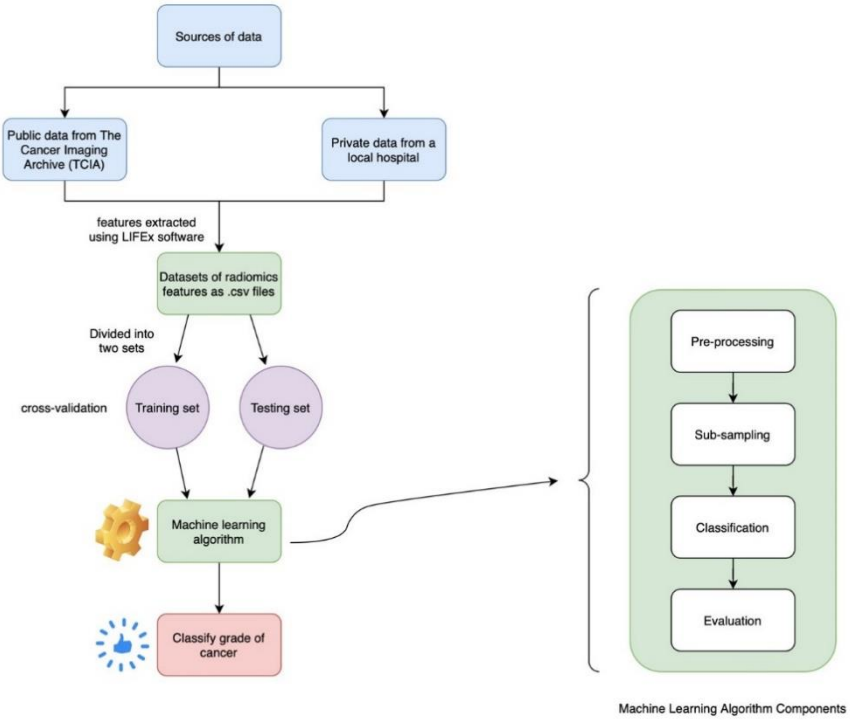

Figure 2.1: Algorithm workflow

2.2 Data Acquisition

Data acquisition is a very vital part of this project as the amount of data used has a direct correlation with the prediction powers of the algorithm. Data is required to train and test the algorithm, check how well it performs. The main dataset used in this project is Diffusion-Weighted MR images from cervical cancer patients. Most of these cervical cancer images were obtained from the largest hospital in Qatar. An issue that was raised is the small amount of data present, therefore a public source was also utilized to obtain these medical images. After obtaining the DWI-MR images the appropriate radiomics features needed to be extracted, this was done through the LifeX program. The output from the program is displayed on an excel

sheet, giving the values for various radiomics features selected previously. The LifeX interface is seen below.

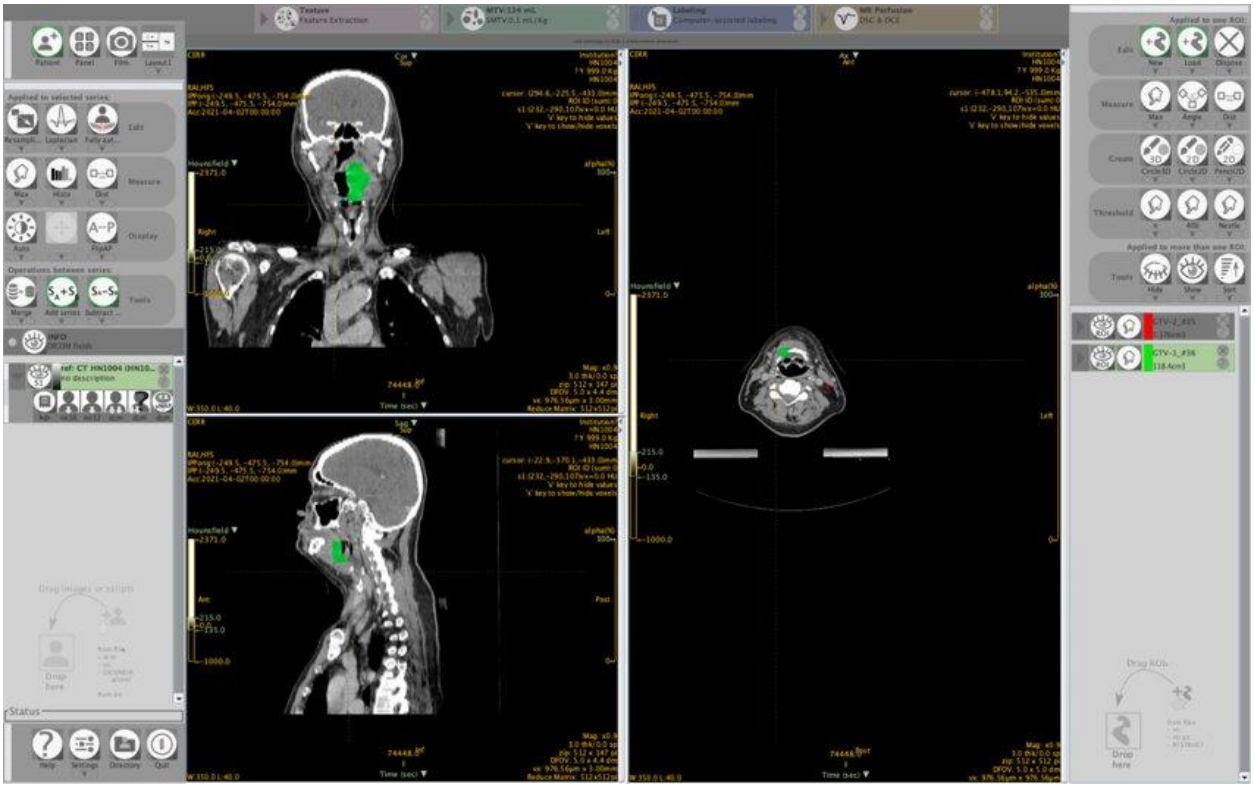

Figure 2.2: Head and neck cancer patient CT image in LifeX software

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	CONVENTIO	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	CONVENTIOI	DISCRETIZED	DISCRETIZED	DISCRETIZED	DISCRETIZED	DISCRETIZED	DISC
2	-123.58233	18.3574326	54.1267022	95.190979	-20.545182	33.9421272	65.1983604	-0.6857406	2.2799293	-0.7200707	0.485	1.021	0	88	102.333713	5.42174626	110	98	
3	-712.56866	61.0091558	35.1496524	618.716003	48.4853535	64.6984634	77.0975075	-3.3099832	62.0516935	59.0516935	0.485	1.021	55413.2462	29	106.599404	3.52675091	162	105	
4	-97.015923	66.7439455	30.9280537	150.741531	50.0398483	76.2613373	88.5277786	-1.138377	4.35554326	1.35554326	0.485	1.021	20.0271606	91	107.177747	3.1010528	116	106	
5	-76.786339	58.5433986	25.0332837	144.104294	45.8295097	61.6727104	76.3613663	-1.0914999	5.23036158	2.23036158	0.485	1.021	366.210938	93	106.350125	2.52309571	115	105	
6	-343.21103	73.3739932	41.0412441	179.172714	68.5380478	79.5064468	88.8678818	-3.8924123	26.5275164	23.5275164	0.485	1.021	803.947449	66	107.835054	4.10931637	118	107	
7	-54.405685	50.6048182	27.3728453	115.042618	37.4926262	54.4859257	70.3340492	-0.9563061	4.04976882	1.04976882	0.485	1.021	0	95	105.568871	2.73887846	112	104	
8	-92.762566	39.0369038	42.965036	111.872803	14.8366933	51.8562279	70.7040596	-0.9795503	3.14820245	0.14820245	0.485	1.021	0	91	104.381496	4.31552887	112	102	
9	-1000	16.6756459	148.977692	306.812622	20.944624	60.287384	75.710495	-4.5472347	26.92035	23.92035	0.485	1.021	6465.91187	1	102.169033	14.8850908	131	103	
10	-1000	63.7370858	133.514642	607.562988	58.6245575	82.0091972	102.896761	-3.9937429	26.7435307	23.7435307	0.485	1.021	3560.06622	1	106.88117	13.355714	161	106	
11	-121.80045	54.1878389	41.9912791	135.922852	39.1307545	68.7128754	83.1162396	-1.5152306	4.98249282	1.98249282	0.485	1.021	295.639038	88	105.918152	4.20934491	114	104	
12	-1000	35.7092467	126.980308	501.323364	32.8345528	55.2336769	73.442009	-4.426519	30.6702263	27.6702263	0.485	1.021	2036.09467	1	104.073536	12.6938299	151	104	
13	-63.676521	49.6281105	39.3173274	148.443222	24.7660093	56.3204136	80.626976	-0.6090798	2.72497974	-0.2750203	0.485	1.021	0	94	105.486456	3.95136024	115	103	
14	-98.92437	50.9866586	45.644953	127.482361	24.1227245	63.4009361	88.5036163	-0.9987665	3.28298748	0.28298748	0.485	1.021	0	91	105.604287	4.58285073	113	103	
15	-1000	37.1018391	156.075232	319.938385	52.3841476	77.1173554	94.1891937	-4.2075741	22.4063427	19.4063427	0.485	1.021	3367.5783	1	104.209515	15.6110061	132	106	
16	-1000	35.9982757	163.301259	839.102539	34.5377798	52.9752903	68.8577099	-2.4925196	20.5167742	17.5167742	0.485	1.021	1783.37097	1	104.113828	16.3296984	184	104	
17	-427.9631	48.7348682	61.6230094	404.516205	25.4217739	54.6544972	77.232687	-1.0800651	10.5709295	7.57092952	0.485	1.021	2471.92383	58	105.368407	6.1711406	141	103	
18	-781.2196	53.2293944	66.3068043	242.587418	55.3351364	68.2790222	77.2344971	-5.4056501	43.515311	40.515311	0.485	1.021	3788.94806	22	105.821838	6.64403094	125	106	
19	-10.331131	80.9660807	23.119445	129.3069	69.2105427	84.3241158	96.6924114	-0.7840648	3.575306	0.575306	0.485	1.021	0	99	108.575529	2.31986156	113	107	
20	-646.09216	83.5655696	66.5461087	254.334625	64.9262848	91.0374451	116.922028	-3.4204192	24.3758981	21.3758981	0.485	1.021	901.222229	36	108.859204	6.65833897	126	107	
21	-97.923935	51.0423923	49.3479104	229.034363	17.078474	62.6525307	90.2661781	-0.5862007	2.7681198	-0.2318802	0.485	1.021	3499.3033	91	105.604976	4.94399037	123	102	
22	-102.99611	69.2435566	39.7785561	144.934586	52.2304039	81.7243767	96.8497543	-1.2136928	3.9699818	0.9699818	0.485	1.021	498.264136	90	107.412059	3.99688768	115	106	
23	-428.80652	38.465325	50.7787678	107	28.7626228	47.8232994	63.7076149	-2.9449498	18.8656538	15.8656538	0.485	1.021	0	58	104.328455	5.08779059	111	103	
24	-955.76386	68.5849297	69.9989046	508.086945	66.0725365	79.0239868	90.8800392	-7.8138222	81.645	78.645	0.485	1.021	1538.70074	5	107.357703	7.00731858	151	107	
25	-373.06015	-1.8615276	87.606681	121.012909	-29.399981	23.0210476	56.8758593	-1.7890031	6.44103915	3.44103915	0.485	1.021	0	63	100.31677	8.75045508	113	98	

Figure 2.3: Generated .csv extracted feature file

2.3 Classifier selection

A lot of emphasis is placed on the predicting powers of the algorithm, therefore various codes were tested to choose the set that produces the most reliable results. As a result, an assortment of classifiers and preprocessing codes were tested. The algorithms tested include a combination of the following pre-processing codes (Principal Component Analysis (PCA) and Synthetic Minority Oversampling Technique (SMOTE)) along with the following classifying algorithms: Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree (DT), X-Gradient Boost (XGB), Light Gradient Boosting Machine (LGBM), CatBoost, Dense Neural Network (DNN), Support Vector Machine (SVM), Naïve Bayes (NB) and Support Vector Classifier Recursive Feature Elimination (SVC RFE). Cross validation was used to find the classification configuration with the optimum performance for each classifier.

Sample size is a vital factor that affects the prediction performance, and so does the balance of the data. Cancer is usually discovered at a later stage, and therefore the cancer data is usually skewed towards the more advanced cancer cases. SMOTE is used to balance the different class distributions by increasing the minority class classes via replication to make a perfectly balanced data set. On the other hand, PCA is used as a form of dimensionality reduction since a large number of features are extracted from medical images, with many of them simply being either noise or highly correlated, ultimately reducing the performance.

The metric of choice used to evaluate the performance of each of the potential algorithm combinations was the Area Under the Curve (AUC). To further increase the performance of the algorithm, an ensemble model was built. An ensemble model averages the different predicted probabilities of each class from the selected classifiers, combining them to form a new

prediction. This method is used in the XGB algorithm and is one of the reasons it is such a powerful classifier.

3. RESULTS

3.1 Classifiers Results

In this study, two different cancer data sets were used, head and neck cancer data was obtained from a public archive while cervical cancer data was obtained provided by our collaborators. The head and neck data contained information from 259 patients, the largest stage percentage was stage 4 with 69% while the lowest stage represented was stage 2 with only 5.4%. For this dataset, both SMOTE and PCA were tested along with the different classifiers mentioned in Methods and the resulting tables can be found below. The evaluating metric of choice for this data set was a combination of precision, recall, and Mathews Correlation Coefficient (MCC). Precision and recall are ratios of confusion matrix elements with a value closer to 1 being best. Precision is an evaluation metric on the relevance of the results, recall is a metric which refers to the percentage of accurately classified and relevant results, and MCC is the evaluation metric which determines the linkage between the observations and predictions with a value of 1 being best. The four tables below (Table 1, Table 2, Table 3 and Table 4) show the results of each classifier used on the head and neck data for stages 1, 2, 3, 4. From the results, we can deduce several observations. First, most classifiers perform very poorly in stage 2, this is because of the quality of the data since stage 2 was the least represented only making up only 5% of the data. Secondly, XGBoost outperforms all the other classifiers which proves its powerful classification abilities. XGBoost is a classifier that is an ensemble of a range of tree-based classifiers optimized with speed and performance as a goal. XGBoost is one of the most popular *classifiers* and tends to excel when it comes to tabular data and it is therefore not a

surprise that it outperforms the other classifiers. PCA was also tested but proved to have an adverse effect on the performance of the classifiers.

Table 1: Results of the various classifiers for stage 1 of head and neck.

Classifier	Precision	Recall	Accuracy	MCC
KNN	0.444	1	0.769	0.399
KNN + SMOTE	0.188	0.750	0.477	0.305
Naïve Bayes	0.250	0.5	0.446	0.286
Naïve Bayes + SMOTE	0.5	0.75	0.554	0.383
Decision Trees	0.364	1	0.677	0.337
Decision Trees + SMOTE	0.15	0.750	0.323	0.152
XGBoost	0.861	0.733	0.861	0.446
SVC (rbf kernel)	Nan	0	0.861	0
Random Forest	0.714	0.555	0.907	0.579
Light GBM	0	0	0.892	-0.039
CatBoost	0.625	0.556	0.892	0.528
DNN	1	0.111	0.635	0.189

Table 2: Results of the various classifiers for stage 2 of head and neck.

Classifier	Precision	Recall	Accuracy	MCC
KNN	0	0	0.769	0.399
KNN + SMOTE	0.111	0.5	0.477	0.305
Naïve Bayes	0.200	0	0.446	0.286
Naïve Bayes + SMOTE	0.2	0	0.554	0.383
Decision Trees	0	0	0.677	0.337
Decision Trees + SMOTE	0	0	0.323	0.152
XGBoost	0.938	0.491	0.938	-0.027
SVC (rbf kernel)	0.046	1	0.046	0
Random Forest	Nan	0	0.953	0
Light GBM	0	0	0.938	-0.027

CatBoost	Nan	0	0.953	0
DNN	0	0	0.635	0.189

Table 3: Results of the various classifiers for stage 3 of head and neck.

Classifier	Precision	Recall	Accuracy	MCC
KNN	1	0.091	0.769	0.399
KNN + SMOTE	0.421	0.727	0.477	0.305
Naïve Bayes	0.276	0.727	0.446	0.286
Naïve Bayes + SMOTE	0.320	0.727	0.554	0.383
Decision Trees	0.417	0.455	0.677	0.337
Decision Trees + SMOTE	0.25	0.364	0.323	0.152
XGBoost	0.892	0.641	0.892	0.306
SVC (rbf kernel)	0.092	1	0.092	0
Random Forest	0.092	1	0.092	0
Light GBM	0	0	0.892	-0.039
CatBoost	Nan	0	0.907	0
DNN	0	0	0.635	0.189

Table 4: Results for the various classifiers for stage 4 of head and neck.

Classifier	Precision	Recall	Accuracy	MCC
KNN	0.818	0.938	0.769	0.399
KNN + SMOTE	0.905	0.396	0.477	0.305
Naïve Bayes	0.944	0.354	0.446	0.286
Naïve Bayes + SMOTE	0.958	0.479	0.554	0.383
Decision Trees	0.833	0.729	0.677	0.337
Decision Trees + SMOTE	0.933	0.292	0.323	0.152
XGBoost	0.646	0.601	0.646	0.189
SVC (rbf kernel)	nan	0	0.276	0
Random Forest	0.773	0.872	0.723	0.237
Light GBM	0	0	0.892	-0.039
CatBoost	0.784	0.851	0.723	0.261
DNN	0.627	1	0.635	0.189

The cervical cancer data was used to classify the cancer grade of the patients. The data set contained 66 patients, 58 % grade 3 and 42% grade 2. The evaluation metric of choice for this binary classification was the AUC, the classifier that achieved the highest AUC score was the LGBM classifier with an AUC score of 0.66. A higher AUC score was achieved however, using an ensemble model which combined the LGBM, RF and XGB classifiers to get an AUC score of 0.70 AUC. The results of each of the classifiers can be seen below in table 5.

Table 5: Results for the various classifiers for cervical cancer grade

Classifiers used	AUC score
LGBM, RF, XGB & KNN	0.5535
LGBM, RF & XGB	0.6964
LGBM, RF & KNN	0.5357
LGBM, XGB & KNN	0.5714
RF, XGB & KNN	0.4821
LGBM & RF	0.6607
LGBM & XGB	0.6428
LGBM & KNN	0.5357
RF & KNN	0.5
RF & XGB	0.5357
KNN & XGB	0.5
KNN	0.4553
RF	0.5714
XGB	0.5446
LGBM	0.6607

4. CONCLUSION

4.1 Conclusion

In this study, an optimal radiomics based model for both the head and neck staging classification and the cervical cancer grading classification were investigated. A variety of different classifiers were tested, XGB proved superior for head and neck staging while an ensemble model containing LGBM, RF and XGB had the best performance for the cervical cancer staging. It was also observed that the unfiltered dimensionality reduction PCA algorithm and the sub-sampling SMOTE technique did not have positive effect when used in conjecture with all the classifiers.

4.2 Future recommendation

Artificial Intelligence is a rapidly evolving field, and cancer research is a never ending one. One of the main constraints we faced in our project is the data availability, for that our recommendation would be to use a data augmentation algorithm or to acquire larger data sets.

An alternative classification method that can be further looked into is transfer learning. Transfer learning is not as limited by data size as other machine learning algorithms and should prove to achieve better results given the availability of the medical imaging.

REFERENCES

[1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortettieulent J, Jemal A. Global cancer statistics, 2012. Ca A Cancer Journal for Clinicians 2015;65:87– 108.

[2] Y. Ai, H. Zhu, C. Xie, and X. Jin, “Radiomics in cervical cancer: Current applications and future potential,” Crit. Rev. Oncol. Hematol., vol. 152, no. May, p. 102985, 2020.

[3] Jones, J., 2020. FIGO Staging System | Radiology Reference Article | Radiopaedia.Org. [online] Radiopaedia.org. Available at: <https://radiopaedia.org/articles/figo-staging-system> [Accessed 25 December 2020].

[4] S. Lee and M. Atri, "2018 FIGO Staging System for Uterine Cervical Cancer: Enter Cross-sectional Imaging", Radiology, vol. 292, no. 1, pp. 15-24, 2019. Available: https://pubs.rsna.org/doi/full/10.1148/radiol.2019190088. [Accessed 26 December 2020].

[5] M. Idris, "Radiomics | Radiology Reference Article | Radiopaedia.org", Radiopaedia.org, 2020. [Online]. Available: https://radiopaedia.org/articles/radiomics. [Accessed: 24- Dec- 2020].

[6] Blazic IM, Lilic GB, Gajic MM. Quantitative assessment of rectal cancer response to neoadjuvant combined chemotherapy and radiation therapy: comparison of three methods of positioning region of interest for ADC measurements at diffusion-weighted MR imaging. Radiology 2017;282:418–428.

[7] U. Schick, F. Lucia, G. Dissaux, D. Visvikis, B. Badic, I. Masson, O. Pradier, V. Bourbonne, and M. Hatt, “MRI-derived radiomics: Methodology and clinical applications in the field of pelvic oncology,” Br. J. Radiol., vol. 92, no. 1104, 2019.

[8] J. Chen, Y. Zhang, B. Liang, and Z. Yang, “The utility of diffusion-weighted MR imaging in cervical cancer,” Eur. J. Radiol., vol. 74, no. 3, pp. e101–e106, 2010.

[9] Y. Liu, Y. Zhang, R. Cheng, S. Liu, F. Qu, X. Yin, Q. Wang, B. Xiao, and Z. Ye, “Radiomics analysis of apparent diffusion coefficient in cervical cancer: A preliminary study on histological grade evaluation,” J. Magn. Reson. Imaging, pp. 1–11, 2018.

[10] S. Leibfarth, R. M. Winter, H. Lyng, D. Zips, and D. Thorwarth, “Potentials and challenges of diffusion-weighted magnetic resonance imaging in radiotherapy,” *Clin. Transl. Radiat. Oncol.*, vol. 13, pp. 29–37, 2018, doi: 10.1016/j.ctro.2018.09.002.

[11] P. Z. McVeigh, A. M. Syed, M. Milosevic, A. Fyles, and M. A. Haider, “Diffusion-weighted MRI in cervical cancer,” *Eur. Radiol.*, vol. 18, no. 5, pp. 1058–1064, 2008, doi: 10.1007/s00330-007-0843-3.

[12] A. LIFEx, C. LIFEx, P. Features, M. us, S. demos and L. available, "LIFEx", *Lifexsoft.org*, 2021. [Online]. Available: https://www.lifexsoft.org/. [Accessed: 13- Jan- 2021].

[13] B. Grabowski, "“P< 0.05” Might Not Mean What You Think: American Statistical Association ClarifiesPValues", *Journal of the National Cancer Institute*, vol. 108, no. 8, p. djw194, 2016. Available: 10.1093/jnci/djw194.

[14] J. M. Winfield *et al.*, “Separation of type and grade in cervical tumours using non-mono-exponential models of diffusion-weighted MRI,” *Eur. Radiol.*, vol. 27, no. 2, pp. 627–636, 2017, doi: 10.1007/s00330-016-4417-0.

[15] L. Oettl, N. Ravi, M. Schneider, M. F. Scheller, M. Mitre, S. Gouveia, R. C. Froemke, M. V Chao, W. S. Young, A. Meyer-lindenberg, V. Grinevich, R. Shusterman, and W. Kelsch, “Radiomics Assessment of Bladder cancer Grade Using Texture Features from Diffusion -Weighted imaging,” vol. 90, no. 3, pp. 609–621, 2017.

[16] "Cancer Grade Vs. Cancer Stage", *MD Anderson Cancer Center*, 2021. [Online]. Available: https://www.mdanderson.org/patients-family/diagnosis-treatment/a-new-diagnosis/cancer-grade-vs--cancer-stage.html. [Accessed: 13- Jan- 2021].

[17] Lin Y, Li H, Chen Z, et al. Correlation of histogram analysis of apparent diffusion coefficient with uterine cervical pathologic finding. AJR Am J Roentgenol 2015;204(5):1125e31

[18] Guan Y., Li W., Jiang Z. Value of whole-lesion apparent diffusion coefficient (ADC) first-order statistics and texture features in clinical staging of cervical cancers. Clin. Radiol. 2017;72:951–958