

Classification of Neuroblastoma Histopathological Images Using Machine Learning

Adhish Panta¹, Matloob Khushi¹✉, Usman Naseem¹, Paul Kennedy³, and
Daniel Catchpoole^{2,3}✉

¹ School of Computer Science, The University of Sydney, Camperdown NSW 2006,
Australia matloob.khushi@sydney.edu.au

² Children’s Cancer Research Unit, The Children’s Hospital at Westmead,
Westmead, NSW, Australia

³ School of Computer Science, UTS, Sydney Australia
daniel.catchpoole@health.nsw.gov.au

Abstract. Neuroblastoma is the most common cancer in young children accounting for over 15 % of deaths in children due to cancer. Identification of the class of neuroblastoma is dependent on histopathological classification performed by pathologists which are considered the gold standard. However, due to the heterogeneous nature of neuroblast tumours, the human eye can miss critical visual features in histopathology. Hence, the use of computer-based models can assist pathologists in classification through mathematical analysis. There is no publicly available dataset containing neuroblastoma histopathological images. So, this study uses dataset gathered from The Tumour Bank at Kids Research at The Children’s Hospital at Westmead, which has been used in previous research. Previous work on this dataset has shown maximum accuracy of 84%. One main issue that previous research fails to address is the class imbalance problem that exists in the dataset as one class represents over 50% of the samples. This study explores a range of feature extraction and data undersampling and over-sampling techniques to improve classification accuracy. Using these methods, this study was able to achieve accuracy of over 90% in the dataset. Moreover, significant improvements observed in this study were in the minority classes where previous work failed to achieve high level of classification accuracy. In doing so, this study shows importance of effective management of available data for any application of machine learning.

1 Introduction

Neuroblastoma is the most common cancer diagnosed in children in the first year of life and accounts for nearly 15% of deaths in children due to cancer[1, 2]. Neuroblast tumours evolve from immature neuroblasts in the sympathetic nervous system during the embryonic, fetal or postnatal stage in children. The disease spreads typically through bone, bone marrow and the liver, and the tumour can

be spotted as mass lesions in areas like neck, chest, abdomen and pelvis[1, 2]. Cellular heterogeneity is one of the distinctive features of neuroblastoma[3]. As a result of this feature, neuroblastic tumours show unexpected clinical behaviour, which includes spontaneous regression and aggressive progression. As such, it is common to see genetic materials achieve gains and loss rapidly when neuroblastoma is evolving.

Pathologists use the Shimada system to classify whether a tumour is favourable or unfavourable, which is considered the gold standard in neuroblastoma classification[3]. The Shimada system considers three key factors, which are: age of the patient, the category of the Neuroblast tumour and the Mitosis-Karyorrhexis index (MKI)[3]. To categorise neuroblast tumours, pathologists examine thin tissues using optical microscopes in different magnifications. While identifying the category of tumours, pathologists use several morphological features such as the presence of neuropil, cellularity, nuclear size and shape[3]. However, due to the complex and heterogeneous nature of neuroblastoma, pathologists can get misleading results. The use of machine learning techniques for feature extraction can reveal information and relationships not visible to the human eye. Moreover, the use of Computer-Aided Design (CAD) systems in the health sector offer benefits such as improvement of the overall speed and quality of the diagnosis process by eliminating human fatigue, acting as a tool for the second opinion and assisting with the shortage of medical experts[3]. These benefits are especially significant for neuroblastoma as the age of a patient at the time of diagnosis is vital for the prognosis outcome.

There is a lack of significant exploration of the use of machine learning techniques in for classification of Neuroblastoma types. One of the recent work was done by S. Gheisari et al. [5–7] who classified neuroblastoma histopathological images into five categories which were: undifferentiated neuroblastoma, ganglioneuroblastoma, ganglioneuroma, poorly-differentiated neuroblastoma, and differentiating neuroblastoma. A range of low level and high-level feature extraction techniques were used in previous research to achieve accuracy of around 84%. From a medical perspective, more accurate results would be desirable to increase confidence and improve the chances of computer-based systems being used to assist experts. Furthermore, [5–7] also identified that there was a high degree of misclassification between poorly-differentiated and differentiating neuroblastoma classes. From a biological perspective, these misclassifications are significant as they can result in patients being overtreated or undertreated. Hence there is keen interest in improving existing methods.

This paper aims to improve the previous work performed by S. Gheisari et al. [5–7], where neuroblastoma images were successfully classified into five categories. The dataset used for this work is the same as used in [5–7] which was gathered from The Tumour Bank at Kids Research at The Children’s Hospital at Westmead. Through the exploration of previously used feature extraction methods and existing data optimisation techniques, this study aims to improve the overall accuracy metrics achieved in previous neuroblastoma research. The contributions made by this study is summarised below:

1. Explored the impact of different feature extraction and machine learning techniques on the performance for classification of neuroblastoma histopathological images.
2. Evaluated the effects of combining outcomes from different feature extraction techniques on performance metrics.
3. Explored the effects of data engineering techniques such as resampling the given dataset on the overall performance.
4. Improved overall accuracy metrics through the use of these techniques

The rest of the paper is as follows: Section 2 discusses the methodology used. Section 3 shows the results. Section 4 presents the discussion and Section 5 provides a conclusion

2 Methodology

To effectively conduct the intended research, this work follows a structure commonly used for the application of machine learning in the medical informatics domain. The first step involves relevant data needs to be collected, or existing dataset needs to be selected to conduct the experiments. Collecting medical image is a complex process which involves taking tissue samples from high resolution microscopes. It would also require expert medical professionals to categorize the data so that models can be trained. So, the collection of raw data was considered out of scope, and an existing neuroblastoma dataset was used to conduct the experiments. The next step involves feature engineering, where the aim is to extract relevant features from available image data. Multiple feature and data engineering techniques are explored in work conducted. Then machine learning classification algorithms are implemented for classification. In this study, SVM classification is used for uniform comparison as a greater focus on the feature extraction and data engineering methods. Finally, the designed models are tested and evaluated using 5-fold cross-validation, accuracy, precision, recall and F-1 score.

2.1 Dataset

The dataset used for this study is the same as used by S. Gheisari et al. in [5–7]. The dataset was gathered from The Tumour Bank at Kids Research at The Children’s Hospital at Westmead and is the most comprehensive available dataset for research in neuroblastoma. There is no publicly available dataset for neuroblastoma research, so the authors granted access to this dataset. The dataset contains 1043 images gathered from stained tissue biopsy slides of 125 patients. The tissue slides were scanned by using a software called Imagescope under 40x magnification. Each image was cropped to include 300 by 300 pixels to provide a balance between achieving a reasonable computational time and preserving critical information in each image. It was also ensured that each cropped image contained areas that best represent each category of neuroblastoma. This is a

slight limitation of the dataset because it does not directly mimic a real-world scenario where data can have noise points. Expert pathologists classified images in the dataset according to the Shimada System where the categories included: poorly differentiated neuroblastoma, differentiating neuroblastoma, undifferentiating neuroblastoma, ganglioneuroma and ganglioneuroblastoma.

As discussed previously, neuroblastoma has a high degree of intra-class variance. The extent of intra-class variance in the available dataset can be seen in Fig.1. Both (a) and (b) in the figure belong to the same class of neuroblastoma (differentiating neuroblastoma). However, as the blue circled section in the image indicates, they have neuroblast cells of differing sizes. This is a prominent feature of neuroblast tumours and has been extensively captured in the dataset. The

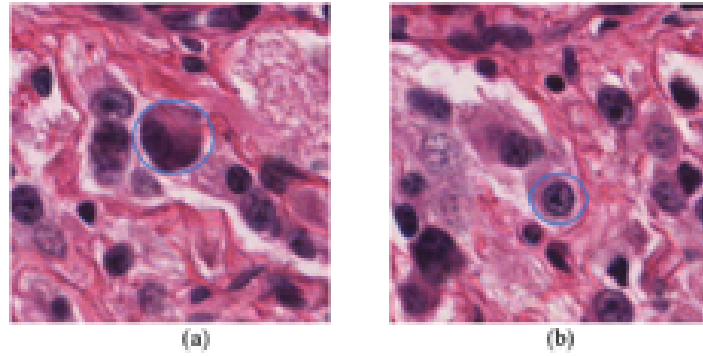


Fig. 1. Example of intra-class variance in neuroblastoma

overall dataset used for the experimental setup can be summarised in Table 1. The table shows data distribution for each class and the number of patients from whom the images were gathered. It can be seen that overall, there are 1043 data samples from 125 patients. It can also be seen that there is a high degree of imbalance in the used dataset. The most common class is poorly-differentiated neuroblastoma as it represents over 50% of the data with 571 samples. For the least common class (ganglioneuroblastoma), there are only 46 samples available which represents less than 5% of the dataset and is gathered from 8 out of 125 patients.

2.2 Experimental Setup

The study explored a range of experimental approaches around feature extraction and data undersampling and oversampling to determine the effects they have in the given dataset. These experimental setups are explained below. To evaluate the setups, k-fold stratified k-fold cross validation was used. A stratified

Table 1. Dataset used in the study

Neuroblastoma Tumour Class	No. Cropped Images	No. Patients
Poorly Differentiated	571	77
Differentiating	187	12
Undifferentiated	155	10
Ganglioneuroma	84	28
Ganglioneuroblastoma	46	8
Total	1043	125

k-fold cross validation ensures that for each split in k, the original distribution of dataset is maintained. Accuracy, precision, recall and F1-score were used as metrics used for comparison of the different approaches and are presented in the results section.

Setup 1 – Scale Invariant Feature Transform(SIFT) + Bag of Visual Words (BOVW) + SVM

SIFT: SIFT feature extractor was introduced by[8] and is used to extract distinctive features that are invariant to scale rotation and illumination. The SIFT method finds the keypoints in an image by executing four key steps: detect scale-space extrema, localise keypoints, assignment orientation and descriptor representation. For any given image, the SIFT algorithm returns two key components. A set of keypoints and a descriptor for each keypoint. SIFT has three key parameters which are the width of the gaussian for scale-space extrema detection, contrast threshold for the elimination of low contrast keypoints and edge threshold for the elimination of edges. These parameters have been tuned according to the recommendations from[7]. The width of the gaussian was set to 1.7; the contrast threshold was set to 0.04 and edge threshold was set to 11.

BOVW: SIFT extracts thousands of feature points from each image, each of which is described by 128 element vectors. BOVW is a commonly used technique to encode features in image processing that is adapted from the Bag of Words algorithm used in Natural Language Processing (NLP)[9]. The BOVW algorithm takes the SIFT features extracted in the previous step and performs clustering over the data. Each cluster identified in this process acts as a visual vocabulary which describes the image. Once the visual vocabulary is established, for each image, a frequency histogram is created to count the occurrence of each feature. Clustering for BOVW is implemented using the k-means clustering algorithm. The number of clusters defines the size of the codebook, and according to the parameter tuning of[7], the cluster size has been set to 500. A visual representation of BOVW can be seen in Figure 2

Classification: For classification, Support Vector Machines(SVM) was used. Other common classifiers such as K-Nearest Kenghbour and Naive Bayes were not considered because[7] had already established that SVM was the optimal

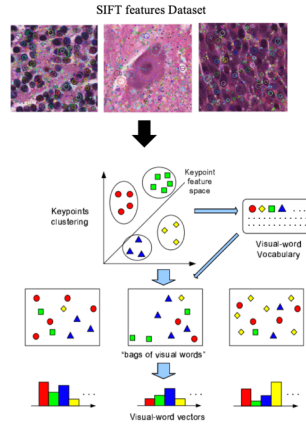


Fig. 2. Illustration of BOVW

classifier for the given dataset. As the study focused on feature extraction and data resampling techniques, exploration of classification algorithms was considered out of scope for this research. So, for all the experimental setups discussed in this study, the SVM model was used. The SVM model used in the study, was setup with a RBF kernel with the kernel parameter set to 0.004. The SVM kernel and parameters were kept consistent for all other experimental setups used in the study

Setup 2: SURF + Bag of Visual Words (BOVW) + SVM

The second experimental setup is similar to the first with the only alteration made to the feature extraction technique used. SURF is used as a feature extractor rather than SIFT in this setup. SIFT is quite effective but is generally more computationally expensive. SURF was developed to improve the speed of SIFT and can be up to three times faster while still providing features invariant to scale, illumination, blur and rotation. H. Bay et al.[10] has also shown that the SURF feature extraction method can outperform SIFT on multiple occasions. SURF has one key parameter – hessian threshold, which is used for keypoint detection. For this study, this parameter was set to 600 as used in[10]. However, citer10,r11 indicate that SURF can outperform SIFT in multiple scenarios. Also, each feature in SURF is described by a 64-dimensional vector as opposed to 128 vector descriptors of SIFT. This means that the BOVW will need parameter tuning. So, the cluster size of 100, 200 and 300 was considered for construction of codebook in BOVW. For classification, SVM classifier with RBF kernel was used.

Setup 3: Combining SIFT and SURF Features

The third experimental setup involves combining SURF and SIFT feature extractors. A similar approach was previously implemented by L. Lenc and P. Král[12] for facial recognition, where results showed that combining SIFT and SURF can outperform state of the art in facial recognition. However, this approach has not been explored in the feature extraction of neuroblastoma tumours. It is expected using this approach of combining keypoint locations and descriptors from two algorithms will provide features that are more robust as compared to using them individually. The process of combining features from SIFT and SURF is as follows: (1) Extract interest points using SIFT extractor; (2) Extract interest points using SURF extractor; (3) Gather descriptors from SIFT extractor; (4) Gather descriptors from SURF extractor (5) Concatenate both descriptors to create a combined descriptor.

Setup 4: Subsampled Features

The experimental setups explored in the previous sections only consider different feature extraction approaches. However, the class imbalance issue remained unanswered. This issue was also not addressed in previous work on neuroblastoma classification by S. Gheisari et al. in[5–7].

Standard techniques used to manage the class imbalance problem are data sampling methods such as oversampling and undersampling[13, 14]. Oversampling through image augmentation is a common approach used in research but is not suitable for this work because SIFT and SURF are robust to scale, blur, rotation and illumination. Hence this setups looked at undersampling to the representation numbers of the majority classes. While this method might appear suitable for the given problem, it has one severe limitation. As discussed in[15] merely removing data samples from the majority class can result in loss of vital information. Thus, the approach of removing image samples from the available dataset would not be a suitable approach for the given problem. Instead, an approach similar to that suggested in [13, 15] appeared more feasible. So, features extracted from the images will be undersampled. This method will ensure that the most relevant features from each class is preserved.

As performed in previous experimental setups, a feature extraction method is first used to extract robust features from all the images in the training set. For this setup SIFT extractor with precisely set parameters (Gaussian width = 1.7, contrast threshold = 0.04 and edge threshold = 11) is used. After computing the feature points and corresponding descriptors, these descriptors are ordered based on the class they represent. For each class, the set of descriptors are then sorted based on relevance. Once vectors are organised by class and sorted, undersampling is performed to reduce the number of feature vectors in the majority classes. To do this, the number of feature vectors in the lowest class is taken as a limit for the number of features for each class. In the case of the dataset used, the minimum number of features would always be gathered from

the ganglioneuroblastoma class. After the undersampling step, as before BOVW and SVM is used. For the BOVW, cluster sizes of 100, 200, and 300 were tested for parameter tuning.

Setup 5: Resampling for classification

The resampling method proposed in the previous setup only addresses the class imbalance for BOVW where clustering was used. It fails to address the class imbalance issue for classification for SVM. Thus, this experimental setup aims to address this imbalance by exploring oversampling and undersampling techniques before performing classification.

To oversample the minority classes for classification, a method known Synthetic Minority Oversampling Technique (SMOTE) is used in this experiment. This method was introduced by N. V. Chawla et al.[16], where additional samples for the minority class is generated synthetically by observing existing examples. SMOTE generates new data for the minority class by considering the nearest neighbours of existing samples. Another technique that can be applied for resampling is known as near-miss. This method is an undersampling approach where instances of the majority class are removed if they are too close to each other. In this method, first distances between all samples in the majority class and the minority class are calculated.

A combined approach of using SMOTE to oversample the minority classes and then using near-miss to undersample on the majority class (poorly differentiated neuroblastoma) is also explored in this experimental setup. First SIFT is used for feature extraction. Then the undersampling method, as discussed in setup 4, is used. The undersampled feature points are then fed through BOVW for feature encoding. After this, oversampling and undersampling methods are used for the available training data and then finally SVM classifier is used for classification

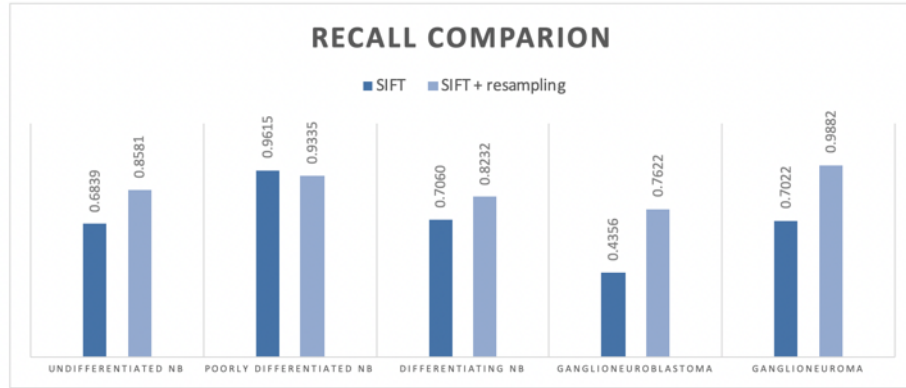
3 Results

This section presents the results achieved with the multiple experimental setups used. Table 2 summaries the key metrics of accuracy, precision, recall and F-1 score achieved from different experimental methods discussed in the previous section. The results table include a summary of the achieved metrics using a stratified 5-fold cross validation. For the sake of comparison, the train-test split in dataset remained consistent throughout the different experimental setups. The results presented in table shows that the final setup generated the best outcome for accuracy, precision, recall F-1 scores. The final setup used the selective undersampling approach for SIFT and and SMOTE for classification. The achieved result, provides evidence that the proposed data manipulation techniques in this paper provided significant improvements in the neuroblastoma dataset. The dramatic improvements in the overall metrics become even more evident when Figure 3 is observed. This figure shows comparison between

Table 2. Results achieved by the experimental setups

Method	Accuracy	Precision	Recall	F-1 Score
Setup 1: Sift for feature extraction	0.8304	0.8573	0.6978	0.7483
Setup 2: SURF + BOVW + SVM	0.7969	0.8081	0.6619	0.7035
Setup 3: Combining SIFT and SURF	0.8477	0.8619	0.7373	0.7801
Setup 4: Subsampled SIFT features	0.8822	0.9151	0.8192	0.8530
Setup 5: Resampling for classification	0.9003	0.9060	0.8730	0.8868

results achieved in previous study and the best results achieved in this study (setup 5) in a class level breakdown. It can be seen that overall, the proposed data sampling approaches provide significant improvements in classification of minority classes such as ganglioneuroblastoma and ganglioneuroma. It also addresses some issues identified in previous work such as the tendency of human experts and machine learning models to misclassify between differentiating and poorly differentiated classes. Overall, the results demonstrate the improvements achieved by this study as compared to previous work on the same dataset.

**Fig. 3.** Class level breakdown and comparison between results achieved with and without sampling approaches

4 Discussion

The work was conducted with the overarching aim of improving classification accuracy in neuroblastoma histopathological images in a given dataset. Due to the complex texture of neuro-blast tumours, machine learning-based classification can be considered beneficial as they can extract features based on mathematical feature extraction techniques. Experimental setups were designed to address the

aims of exploring different feature extraction techniques and analysing the impacts of data engineering techniques such as oversampling and undersampling.

Both SIFT and SURF extract features that are robust to these properties, so their performance were evaluated. Comparing, the results achieved for SIFT and SURF, it was evident that SURF performed poorly compared to SIFT for all the classes. While both SIFT and SURF extract features that are robust to features like, scale and rotation, the mathematical approach taken in these approaches are significantly different. While SIFT uses convolutions of Difference of Gaussian to determine the scale space, SURF uses Laplacian of Gaussian approach with boxed filters to approximate the Difference of Gaussians for scale space detection. As a result of this, features extracted by SIFT are more robust to scale variations compared to SURF. Thus, SURF performed poorly because there is high intra-class scale variance in neuroblast tumours which SIFT was more accurately able to capture.

A combined SIFT and SURF feature extraction method was used to evaluate the effects of combining different feature extraction techniques. Results showed that this approach provided only a slight improvement in overall classification performance. The approach only achieved an improvement of 1.5% as the accuracy of 84.5% was recorded. The only class affected was the undifferentiated neuroblastoma class which saw an improvement of around 7%.

The results achieved by SURF and a combination of SIFT and SURF provided very negligible improvements in performance. The improvements were minimal because these methods did not address the class imbalance issue existing in the dataset. Thus, the class imbalance issue was addressed by using both under-sampling and oversampling approaches in the experimental setups as described elsewhere [17, 18]. By addressing the imbalance issue through these methods, significant improvements in the results were observed. More specifically the overall accuracy improved to over 90% as opposed to the 83% accuracy achieved in the initial model.

Results achieved when the dataset was resampled showed that the applied techniques create a more robust model which can classify all classes with higher accuracy (recall). This improved performance compared to the previous model can be seen in Figure 2 which provides a comparison between initially replicated work and the resampling method developed in this work. In the figure, it can also be seen that the two most improved classes are ganglioneuroblastoma and ganglioneuroma classes where recall score increased by over 30%. Not surprisingly, these two categories were also the two least common class in the dataset. The use of sampling techniques did reduce the accuracy of the poorly differentiated class by 3%, but this is a worthy tradeoff as it is preferable to have higher accuracy overall classes than for just one class. Differentiating neuroblastoma was also identified as a problematic class in previous research because expert pathologists struggled with classification between poorly differentiated and differentiating class. However, the current approach was able to improve the recall of differentiating neuroblastoma class by over 12%. There is also significant medical significance of these improvements in addition to the statistical significance.

Treatment plans for patients are dependent on the category of neuroblastoma they have. So, a more accurate model translates to lower possibilities of patients receiving inadequate treatment.

While the study improves on previous work there are few limitations that can be addressed for future work. This work only looked at the classification of neuroblastoma images using low-level feature extraction combined with a range of data sampling techniques. While results achieved show useful improvements to existing methods, the use of high-level feature extraction techniques can provide even more significant improvements. Use of techniques such as deep neural networks have the potential to extract even more robust and useful features and need to be explored thoroughly in the future. Furthermore, the work only focuses on the classification of patched 300 x 300 stained tissue images into different categories of neuroblastoma. Additionally, the images are carefully curated by expert pathologists and only contain areas that best represent each class. In, a practical scenario, pathologists examine whole tissue images under different magnifications. So, the proposed techniques need to be evaluated with whole slide tissue images under different magnifications.

5 Conclusion

This study began with the objective of improving existing approaches in the histopathological classification of early childhood tumour called neuroblastoma. To improve the results, the study looked at two main approaches. The first approach was to explore alternative low-level feature extraction techniques which could extract more relevant features and improve accuracy. However, this approach did not provide any significant improvements to existing approaches in neuroblastoma. This was because the main problem associated with the dataset was the class imbalance problem. To address the class imbalance in the dataset, both undersampling and oversampling approaches were thoroughly explored. The use of these approaches improved the accuracy of the models to over 90% in the given dataset. More importantly, these methods improved the classification of the minority classes by up to 35%. Such improvements have a high significance in the medical domain because the classification of the neuroblastoma types determine the treatments that patients receive. Pathologist classification using the Shimada system is still considered the gold standard in neuroblastoma and this work only intends to show role computer-based approaches can have to support decision making and act a tool for second reference. Further research is required to classify whole tissue images under different magnifications and predict patient outcomes for practical applications of computer-based approaches in neuroblastoma.

References

1. J. R. Park, A. Eggert, and H. Caron, "Neuroblastoma: biology, prognosis, and treatment," *Pediatric clinics of North America*, vol. 55, no. 1, pp. 97-120, 2008.

2. J. M. Maris, "Recent advances in neuroblastoma," *New England Journal of Medicine*, vol. 362, no. 23, pp. 2202-2211, 2010.
3. H. Shimada et al., "The international neuroblastoma pathology classification (the Shimada system)," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 86, no. 2, pp. 364-372, 1999.
4. J. M. Maris and K. K. Matthay, "Molecular biology of neuroblastoma," *Journal of clinical oncology*, vol. 17, no. 7, pp. 2264-2264, 1999.
5. S. Gheisari, D. R. Catchpoole, A. Charlton, and P. J. Kennedy, "Patched completed local binary pattern is an effective method for neuroblastoma histological image classification," in *Australasian Conference on Data Mining, 2017: Springer*, pp. 57-71.
6. S. Gheisari, D. R. Catchpoole, A. Charlton, and P. J. Kennedy, "Convolutional deep belief network with feature encoding for classification of neuroblastoma histological images," *Journal of pathology informatics*, vol. 9, 2018.
7. S. Gheisari, D. Catchpoole, A. Charlton, Z. Melegh, E. Gradhand, and P. Kennedy, "Computer Aided Classification of Neuroblastoma Histological Images Using Scale Invariant Feature Transform with Feature Encoding," *Diagnostics*, vol. 8, no. 3, p. 56, 2018.
8. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
9. J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval, 2007: ACM*, pp. 197-206.
10. H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision, 2006: Springer*, pp. 404-417.
11. P. Panchal, S. Panchal, and S. Shah, "A comparison of SIFT and SURF," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 2, pp. 323-327, 2013.
12. L. Lenc and P. Král, "A combined SIFT/SURF descriptor for automatic face recognition," in *Sixth International Conference on Machine Vision (ICMV 2013), 2013*, vol. 9067: International Society for Optics and Photonics, p. 90672C.
13. N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 1-6, 2004.
14. D. Ramyachitra and P. Manikandan, "Imbalanced dataset classification and solutions: a review," *International Journal of Computing and Business Research (IJCBR)*, vol. 5, no. 4, 2014.
15. S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," *GESTS International Transactions on Computer Science and Engineering*, vol. 30, no. 1, pp. 25-36, 2006.
16. N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 1-6, 2004.
17. Barlow, H.; Mao, S.; Khushi, M. Predicting High-Risk Prostate Cancer Using Machine Learning Methods. *Data* 2019, 4, 129.
18. Khushi M, Clarke CL, Graham JD. 2014. Bioinformatic analysis of cis-regulatory interactions between progesterone and estrogen receptors in breast cancer. *PeerJ* 2:e654 <https://doi.org/10.7717/peerj.654>