A Modern Approach to Osteosarcoma Tumor Identification Through Integration of FP-Growth, Transfer Learning and Stacking Model

John Sanmartín(✉), Paulina Azuero, and Remigio Hurtado

Universidad Politécnica Salesiana, Cuenca, Ecuador
{jsanmartinq1,pazuero}@est.ups.edu.ec, rhurtadoo@ups.edu.ec
https://www.ups.edu.ec

Abstract. The early detection of cancer through radiographs is crucial for identifying indicative signs of its presence or status. However, the analysis of histological images of osteosarcoma faces significant challenges due to discrepancies among pathologists, intra-class variations, inter-class similarities, complex contexts, and data noise. In this article, we present a novel deep learning method that helps address these issues. The architecture of our model consists of the following phases: 1) Dataset construction: advanced image processing techniques such as dimensionality reduction, identification of frequent patterns through unsupervised learning (FP-Growth), and data augmentation are applied in this phase. 2) Stacking model: we apply a stacking model that combines the strengths of two models: convolutional neural networks (CNN) with transfer learning, allowing us to leverage pre-trained knowledge from related datasets, and a Random Forest (RF) model to enhance the classification and diagnosis of osteosarcoma images. The models were trained on a dataset of publicly available images from The Cancer Imaging Archive (TCIA) [12]. The accuracy of our models is evaluated using classification metrics such as Accuracy, F1 Score, Precision, and Recall. This work provides a solid foundation for ongoing innovation in histology and the potential to apply and adapt this approach to broader clinical challenges in the future.

Keywords: Deep learning · Data Science · Transfer Learning · Ensemble Models · Frequent Patterns · Osteosarcoma

1 Introduction

Osteosarcoma is considered one of the most common primary bone cancers, characterized by its rapid growth and tendency to spread to other areas of the body. It usually affects long bones. Initial symptoms may include persistent bone pain, swelling, and occasionally the presence of a mass or lump in the affected area. According to data from the American Cancer Society (ACS), osteosarcoma accounts for 2 percent of all cancers in children aged 0 to 14 years and 3 percent

Á. Rocha et al. (Eds.): ICITS 2024, LNNS 932, pp. 298–307, 2024.
https://doi.org/10.1007/978-3-031-54235-0_28

of all cancers in adolescents aged 15 to 19 years. Due to its aggressive nature, early diagnosis is crucial for the prognosis and effective treatment of this cancer. However, like other types of cancer, its analysis is complex due to the difficulties associated with the interpretation of histological images and, on the other hand, the lack of communication between data analysts and physicians [8]. Therefore, the use of deep learning to support medical decision-making has been studied through various methods. Additionally, to achieve more effective machine learning, the use of unsupervised learning techniques is recommended to enhance this process [7]. In our method, we propose the use of the **FP-Growth** algorithm: a data mining tool capable of discovering frequent patterns in images. We also employ a **Stacking** modeling approach: combining predictions from multiple base models into a final model, improving overall accuracy by integrating different approaches to the base models and reducing the risk of overfitting. In our approach, we combine the best of two machine learning techniques: **Convolutional Neural Networks (CNNs)**, which extract important features from osteosarcoma images, along with the transfer learning technique that allows us to use pre-trained models with large datasets and adapt them to our needs. **Random Forest (RF)** operates by building multiple decision trees during training and merging them to obtain more accurate and stable predictions. Its ability to handle large datasets with high-dimensional input variables makes it particularly useful in a variety of applications, including the classification of medical images, as in the case of osteosarcoma. Additionally, we experiment with another well-known machine learning technique, the **K-nearest neighbors (KNN)**. It is a technique with good results. However, it is not scalable unless there is prior reduction [2]. Therefore, to test this method, we applied 2 reduction techniques and compared the results obtained for each case. One of them is Principal Component Analysis (PCA) and umap-Learn. In our work, we address several key aspects that contribute significantly to solving issues with the identification of histological images of osteosarcoma. We highlight the most relevant aspects of the proposed work:

- **Implementation of unsupervised learning for identifying frequent patterns using FP-Growth applied to histological images**.
- **Optimized transfer learning for histological data**.
- **Model stacking integrating convolutional neural networks with transfer learning and Random Forest**.
- **Creation and evaluation of multiple machine learning models for the classification of osteosarcoma images using standard quality measures**.

The rest of the content of this article is organized as follows: in Sect. 2, we discuss the relevant context of previous research. Section 3 provides a detailed description of the methodology used, while Sect. 4 presents the design of the experiments conducted. The results and their discussion are presented in Sect. 5. Finally, in Sect. 6, we present the conclusions obtained and outline the way forward for future research.

2 Related Work

The application of deep learning techniques, unsupervised learning, and other methodologies in the diagnosis and classification of pathological conditions through medical images has been extensively researched. In particular, the FP-Growth algorithm has been successfully applied by Agrawal [1] to extract relevant patterns in tomographic data obtained through Computed Tomography (CT) and has identified useful features for the classification of tumors in different organs. Similarly, transfer learning has proven effective in the classification of medical images, as demonstrated by Li et al. [10], who used transfer learning from pre-trained models on large image datasets to enhance accuracy in classifying magnetic resonance imaging (MRI) images of brains with tumors.

To address the challenge of improving classification accuracy, the Stacking approach has gained traction. Wolpert's work [15] addresses the concept of Stacked Generalization (Stacking) as a strategy to combine base models and enhance generalization. Additionally, dimensionality reduction techniques such as PCA and UMAP have been widely used. In the work of Oliveira et al. [6], PCA was applied to reduce the dimensionality of features extracted from melanoma images, improving classification among different types of skin lesions.

Our method merges FP-Growth, transfer learning, and stacking techniques, as well as experimenting with models created from dimensionality reduction techniques to classify osteosarcoma tissues. Additionally, our method aims to surpass the accuracy achieved by an approach previously proposed by Mishra [11]. In their study, Mishra addressed our same case and achieved promising results by implementing a Convolutional Neural Network (CNN) for the classification of osteosarcoma images. With our approach, we aim to build upon this foundation and further advance the improvement of accuracy and effectiveness in classification, contributing to the advancement of artificial intelligence and medicine.

3 Proposed Method

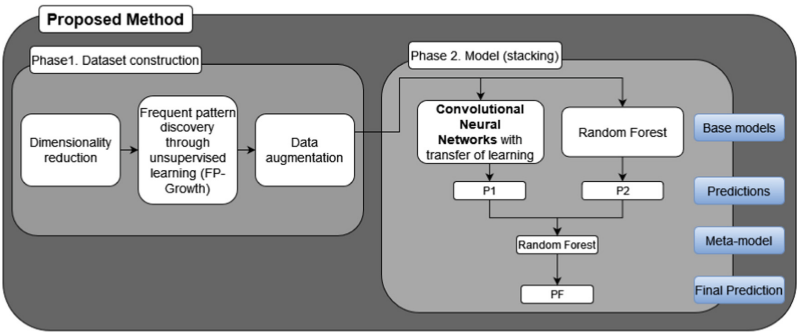

Fig. 1. Proposed Method.

In this section, we present the architecture of the proposed method, as illustrated in Fig. 1.

Our method is divided into two main phases: Phase 1, focused on the construction of the dataset, and Phase 2, focused on the creation of the model using the stacking technique. Each of these phases is further divided into additional sub-phases which we describe below.

3.1 Phase 1 - Dataset Construction

In this phase, we concentrate on preparing image data. This includes dimensionality reduction, enhancing relevant image features using the FP-Growth algorithm, and generating additional data to improve dataset quality.

Subphase 1.1 - Data Loading and Dimensionality Reduction. The original images, possess a size of $1024 \times 1024 \times 3$, which we reduce to $224 \times 224 \times 3$. This allows us to feed the pre-trained model for transfer learning, in addition to reducing computational complexity and processing time.

Subphase 1.2 - Application of the FP-Growth. In this section a list of transactions is created, containing the label (class) associated to each image in the dataset, then transformed into a binary matrix in which each row represents a transaction and each column represents a unique label present in the dataset. By applying the algorithm, relevant combinations of visual features of the images are obtained.

Subphase 1.3 - Data Augmentation. This process is carried out with the frequent patterns obtained by creating a blank mask on which the regions of interest corresponding to the frequent patterns detected in the images are overlaid and highlighted. Then, contrast filters and negative effect are applied to the images, creating two additional variants. Then, rotation, horizontal flipping, brightness adjustment, contrast adjustment and zoom are applied to the images. This improves the generalizability of the model, making it more robust and adaptable to diverse conditions and scenarios.

3.2 Phase 2 – Model (Stacking)

In Phase 2, we focus on the design and development of base models to apply the stacking process, and the results obtained when evaluated are presented in Table 4.

Subphase 2.1 - CNN Model with Transfer Learning. Transfer learning is a technique that allows leveraging the knowledge acquired from pre-trained models and applying it to specific problems. In our case, we implement the DenseNet121 model, which is well-known and widely used in image processing

and classification tasks. Once the model is loaded, we adjust its weights and parameters to adapt it to our specific problem. The parameters used for this model can be found in Table 2. Then, the model is trained using our dataset and predictions are made.

Subphase 2.2 - Random Forest Model. The RF model is created, which is an ensemble learning algorithm that combines multiple decision trees to improve classification accuracy. This process follows a similar structure to the one described earlier, including model creation, training, and finally, making predictions.

Subphase 2.3 - Model Stacking. In this crucial stage, we leverage the advantages of the individual CNN and RF models by creating a meta-model. The central goal of the stacking process is to strategically combine the predictions of these base models to enhance the accuracy of the final classification. The meta-model, in this case, is another Random Forest (RF) model, which is trained using the predictions generated earlier by the CNN and RF models. This enables the meta-model to learn how to optimally weigh and merge the contributions of the base models, capitalizing on their individual strengths to achieve a highly accurate final classification.

4 Experimental Design

In this section, we present the characteristics of the selected dataset (see Table 1), optimization parameters (see Table 2), as well as the metrics used to evaluate the models. The formulas for these quality metrics are presented below.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F1-Score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP, TN, FP, and FN are the numbers of True Positives, True Negatives, False Positives, and False Negatives, respectively.

Table 1. Dataset Description

Dataset	Features	# Samples	# Outputs
Dataset original: Osteosarcoma data from UT Southwestern/UT Dallas for Viable and Necrotic Tumor Assessment (Osteosarcoma Tumor Assessment) [12]	Images of size: $1024 \times 1024 \times 3$	1144	Necrosis, Non-Tumor, Non-Viable-Tumor, Viable
Processed dataset: dimensionality reduction and data augmentation are applied	Images of size $224 \times 224 \times 3$	3432	Necrosis, Non-Tumor, Non-Viable-Tumor, Viable

The table below provides details on the parameters and architecture of each base model used in the stacking. The best results of each model are underlined for differentiation.

Table 2. Parameters of prediction model

Method	Parameters
Convolutional Neural Network (CNN) with Transfer Learning	Pre-trained DenseNet121 Model epochs:1,3,5,**7**,10,15,18,20 batch size:20,30,**35**,40,50,80,90,100 **Architecture:** DenseNet121 Base Model Pre-trained on ImageNet Input Shape: (224, 224, 3), Global Average Pooling: Yes Dense Layer: Units: 256, Activation: ReLU Dropout: 0.5, Output Layer Units: 4 Activation: Softmax (for multiclass classification)
Random Forest (RF)	estimators: 20,40,60,80,**100,**200,300; maximum depth: default
Final meta-model (RF) when applying stacking	Combination of model predictions: **CNN and RF**, estimators: **10**, random state **42**

In addition to the proposed method, experiments were conducted with other models to compare the results. These additional models and methods include a CNN without transfer learning and three KNN models. The first KNN model was fed with the results from the CNN without transfer learning, while the following two KNN models were created using dimensionality reduction techniques, specifically Umap-Learn and PCA. These latter two models were developed with the intention of providing an option for future work that may face extensive datasets. The details of these additional models are presented in the following Table 3.

Table 3. Parameters of other prediction models

Method	Parameters
Convolutional Neural Network (CNN)	epochs:1,3,5,**7**,10,15,18,20 batch size:20,30,**35**,40,50,80,90,100 **Architecture:** **Convolutional 1:** Filters: 8, Kernel Size: (3, 3), Activation: ReLU, Input Size: (224, 224, 3), Max Pooling 1, Window Size: (2, 2) **Convolutional 2:** Filters: 16, Kernel Size: (3, 3) Activation: ReLU Max Pooling 2, Window Size: (2, 2) Flatten Dense 1 Units: 8, Activation: ReLU Dense 2 (Output) Units: 4, Activation: Softmax (for multiclass classification)
Baseline (CNN) [11]	the detailed description of the architecture can be found at [11]
K-Nearest Neighbors (KNN) fed by CNN	neighbors: 5,7,10,9, 11,**15**; metrics: **Euclidean**, Manhattan
KNN with UMAP reduction	neighbors: 5,7,10,9, 11,**15**; metrics: **Euclidean**, Manhattan
KNN with PCA reduction	neighbors: 5,7,10,9, 11,**15**; metrics: **Euclidean**, Manhattan

5 Results and Discussion

The results of the evaluation of the base models for stacking are presented in Table 4. On the other hand, the results of the test models for comparison are shown in Table 5. It is worth noting that all models have been assessed using metrics designed for classification problems.

Table 4. Results of the evaluation of prediction models

Method	Accuracy	Precision	Recall	F1-Score
CNN from Transfer Learning	**0.6593**	**0.6620**	**0.6593**	**0.6126**
Random Forest	**0.7423**	**0.8712**	**0.7423**	**0.7822**
Final meta-model (RF) when applying stacking	**0.9810**	**0.98**	**0.9810**	**0.98**

Table 5. Results of the evaluation of other method

Method	Accuracy	Precision	Recall	F1-Score
Convolutional Neural Network (CNN)	0.4876	0.3234	0.4876	0.3320
Baseline (CNN) [11]	0.84	0.89	0.84	0.86
KNN	**0.4890**	**0.5097**	**0.4890**	**0.4287**
KNN Model fed by CNN	**0.6317**	**0.7001**	**0.6317**	**0.6632**
KNN with PCA	**0.5225**	**0.4854**	**0.5225**	**0.5015**
KNN with Umap-Learn	**0.6011**	**0.5759**	**0.6011**	**0.5849**

As can be observed, our proposed method not only significantly outperforms the CNN approach [11], but also demonstrates superior performance compared to traditional methods. Furthermore, the inclusion of techniques such as PCA and UMAP in our experiments presents an interesting perspective for future researchers, who may consider these techniques as valuable options for addressing extensive datasets or as a starting point for further investigations.

6 Conclusions

In this article, we have proposed and evaluated a modern and effective method to enhance the accuracy in the identification of osteosarcoma tumors. Our approach addresses common challenges in histology, such as discrepancies among pathologists and intracategory variations. To achieve this, we integrated unsupervised learning methods, such as FP-Growth, to highlight frequent patterns, and leveraged transfer learning. This allowed us to develop a CNN model that served as the foundation for our stacking process. We also created a Random Forest model, known for its effectiveness in classification problems. Finally, we developed a stacking model that combines the strengths of the base models, resulting in a significant improvement in classification accuracy. Our results show an overall accuracy of 0.98, demonstrating the effectiveness of our approach in classifying histological images of osteosarcoma.

In addition to achieving high performance, our approach has addressed common challenges in histology, including differences in interpretation among pathologists and intracategory variations. By highlighting frequent patterns and utilizing pre-trained knowledge, we have shown how transfer learning can enhance our models' ability to handle complex histological data. Regarding the stacking model, its capacity to combine the strengths of multiple models has led to a significant improvement in classification accuracy. This underscores the importance of exploring ensemble approaches to tackle critical medical issues, as in our case study. Despite the achievements in this research, there are still opportunities to enhance and expand our approach. We propose the following directions:

- Improvement of training efficiency: Exploring techniques to accelerate model training, such as the use of specialized hardware or model optimization techniques.
- Expansion of the dataset: Collecting and utilizing larger and more diversified datasets to enhance the robustness and generalization of our models.

We believe that these future directions will continue to drive innovation in histology and promote the application of our approach to broader clinical challenges in the future.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
2. Bermejo Casla, G.: Diseño de un algoritmo KNN aplicado a la detección de cáncer cerebral mediante imágenes hiperespectrales. ETSIS_Telecomunicacion (2017)
3. Briones-Velázquez, E.A., Luna-de-la-Luz, C.: Recidiva de osteosarcoma: detección por gammagrafía y SPECT/CT empleando 99mTc-sestamibi. Revista mexicana de pediatría **87**(1), 38–40 (2020)
4. Cánovas-García, F., Alonso-Sarría, F., Gomariz-Castillo, F.: Modificación del algoritmo Random Forest para su empleo en clasificación de imágenes de teledetección. In Aplicaciones de las Tecnologías de la Información Geográfica (TIG) para el desarrollo económico sostenible XVII Congreso Nacional de Tecnologías de Información Geográfica (2016)
5. D'Acunto, M., Martinelli, M., Moroni, D.: Deep learning approach to human osteosarcoma cell detection and classification. In: Choroś, K., Kopel, M., Kukla, E., Siemiński, A. (eds.) MISSI 2018. AISC, vol. 833, pp. 353–361. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-98678-4_36
6. de Oliveira, G.A.P., de Siqueira, M.E.P.B., Vasconcelos, A.C.A., da Silva Torres, R.: Melanoma recognition in dermatoscopy images through deep learning. In: International Workshop on Artificial Neural Networks and Machine Learning, pp. 98–107. Springer, Cham (2017)
7. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 1–12. ACM (2000)
8. Hurtado, R., Guzmán, S., Muñoz, A.: An architecture and a new deep learning method for head and neck cancer prognosis by analyzing serial positron emission tomography images. In: Naiouf, M., Rucci, E., Chichizola, F., De Giusti, L. (eds.) JCC-BD&ET 2023, pp. 129–140. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-40942-4_10
9. Krishnaraj, N., Vidhya, R., Vigneshwar, M., Gayathri, K., Haseena Begam, K., Kavi Sindhuja, R.M.: Pneumonia prediction on X-ray images using CNN with transfer learning. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) ICIPCN 2022, pp. 816–825. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-12413-6_64
10. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from CT volumes. IEEE Trans. Med. Imaging **37**(12), 2663–2674 (2018)

11. Mishra, R., Daescu, O., Leavey, P., Rakheja, D., Sengupta, A.: Histopathological diagnosis for viable and non-viable tumor prediction for osteosarcoma using convolutional neural network. In: Cai, Z., Daescu, O., Li, M. (eds.) ISBRA 2017. LNCS, vol. 10330, pp. 12–23. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59575-7_2
12. Leavey, P., et al.: Osteosarcoma data from UT Southwestern/UT Dallas for Viable and Necrotic Tumor Assessment [Conjunto de datos]. The Cancer Imaging Archive (2019). https://doi.org/10.7937/tcia.2019.bvhjhdas
13. Pannu, A.: Artificial intelligence and its application in different areas. Int. J. Eng. Innovative Technol. **4**, 79–84 (2015)
14. Siegel, R.L., Miller, K.D., Wagle, N.S., Jemal, A.: Cancer statistics, 2023. CA: A Cancer J. Clin. **73**(1), 17–48 (2023). https://doi.org/10.3322/caac.21763
15. Wolpert, D.H.: Stacked generalization. Neural Networks **5**(2), 241–259 (1992)