Search all The European Physical Journal Special Topics articles →



Volume 234, Issue 15

October 2025

Artificial Intelligence and Complex Networks meet Natural Sciences

Favorite Export

JCR Year 2024

European Physical Journal-Special Topics

View title change

ISSN
1951-6355

EISSN
1951-6401

JCR ABBREVIATION
EUR PHYS J-SPEC TOP

ISO ABBREVIATION
Eur. Phys. J.-Spec. Top.

Journal information

EDITION
Science Citation Index Expanded (SCIE)

CATEGORY
PHYSICS, MULTIDISCIPLINARY

LANGUAGES	REGION	1ST ELECTRONIC JCR YEAR
English	FRANCE	2007

Publisher information

PUBLISHER	ADDRESS	PUBLICATION FREQUENCY
SPRINGER HEIDELBERG	TIERGARTENSTRASSE 17, D-69121 HEIDELBERG, GERMANY	24 issues/year

Journal's performance

Journal Impact Factor

The Journal Impact Factor (JIF) is a journal-level metric calculated from data indexed in the Web of Science Core Collection. It should be used with careful attention to the many factors that influence citation rates, such as the volume of publication and citations characteristics of the subject area and type of journal. The Journal Impact Factor can complement expert opinion and informed peer review. In the case of academic evaluation for tenure, it is inappropriate to use a journal-level metric as a proxy measure for individual researchers, institutions, or articles. Learn more

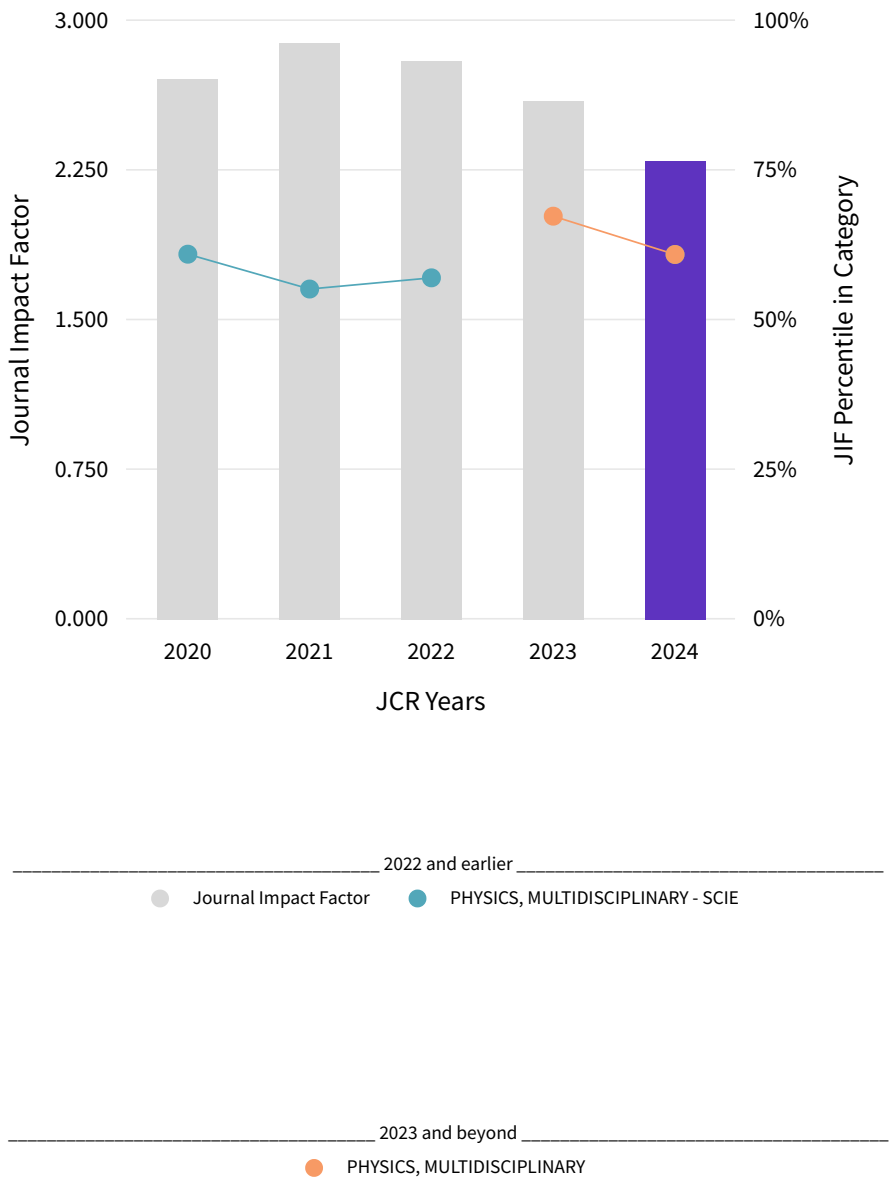
2024 JOURNAL IMPACT FACTOR	JOURNAL IMPACT FACTOR WITHOUT SELF CITATIONS
2.3	2.1
View calculation	View calculation

Journal Impact Factor Trend 2024 Export

Journal Impact Factor contributing items Export

Citable items (615) Citing Sources (496)

TITLE	CITATION COUNT
Artificial neural networks: a practical review of applications involving fractio…	41



A review on vibration energy harvesting technologies: analysis and technologies	20
Dynamical analysis, infections in plants, and preventive policies utilizing the…	20
Steady-state crystallization with a mushy layer: a test of theory with experiments	20
Magneto-hydrothermal performance of hybrid nanofluid flow through a non-…	18
Dynamics and chimera state in a neural network with discrete memristor coupling	16
Optimal synchronization of fractal-fractional differentials on chaotic…	16
Paper based microfluidic devices: a review of fabrication techniques and…	16
Exact solutions for MHD axisymmetric hybrid nanofluid flow and heat transfer…	14
Infinitely many coexisting hidden attractors in a new hyperbolic-type…	14

View All in Web of Science

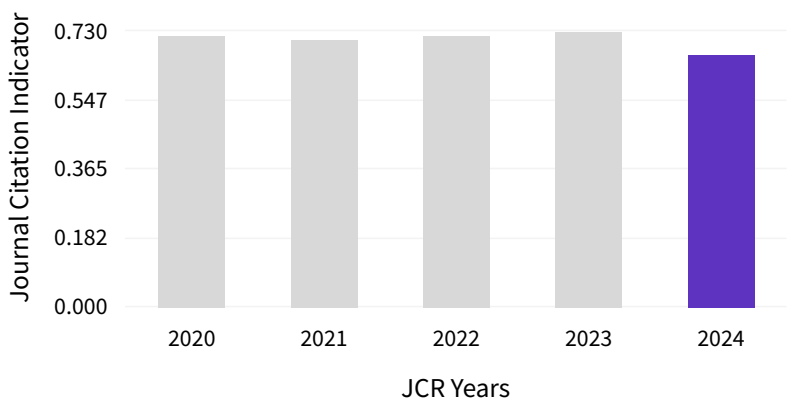
View all years

Journal Citation Indicator (JCI)

Export

0.67

The Journal Citation Indicator (JCI) is the average Category Normalized Citation Impact (CNCI) of citable items (articles & reviews) published by a journal over a recent three year period. The average JCI in a category is 1. Journals with a JCI of 1.5 have 50% more citation impact than the average in that category. It may be used alongside other metrics to help you evaluate journals. Learn more



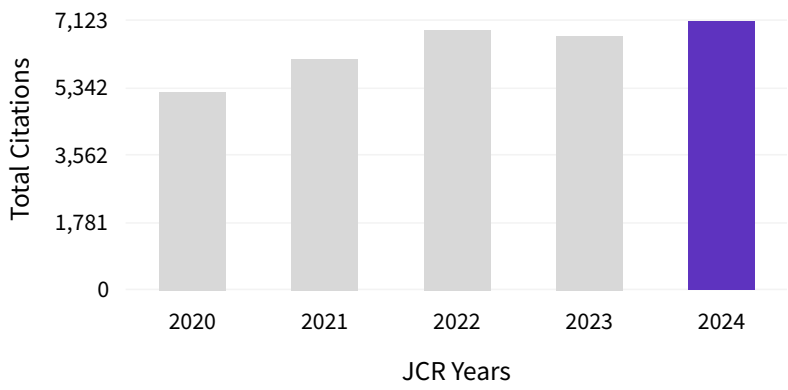
View all years

Total Citations

Export

7,123

The total number of times that a journal has been cited by all journals included in the database in the JCR year. Citations to journals listed in JCR are compiled annually from the JCR years combined database, regardless of which JCR edition lists the journal.



View all years

Citation distribution

Export

The Citation Distribution shows the frequency with which items published in the year or two years prior were cited in the JCR data year (i.e., the component of the calculation of the JIF). The graph has similar functionality as the JIF Trend graph, including hover-over data descriptions for each data point, and an interactive legend where each data element's legend can be used as a toggle. You can view Articles, Reviews, or Non-Citable (other) items to the JIF numerator. Learn more

ARTICLE CITATION MEDIAN

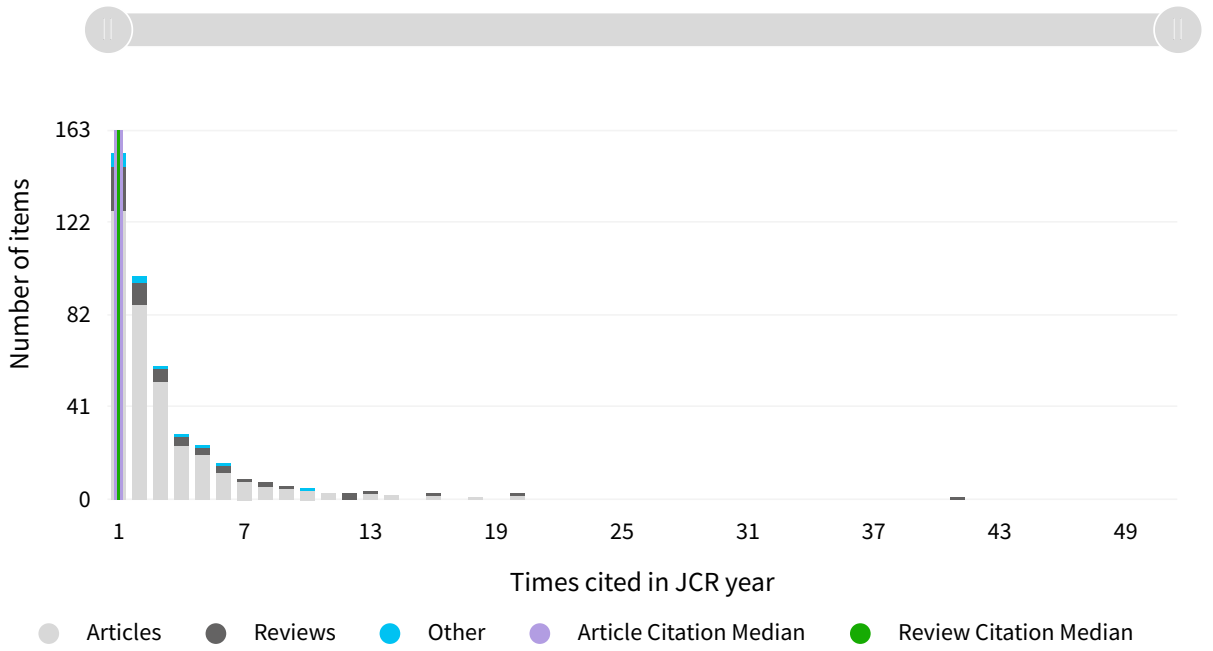
1

REVIEW CITATION MEDIAN

1

UNLINKED CITATIONS

36



TIMES CITED

0

ARTICLES

180

REVIEWS

21

OTHER

42

Open Access (OA)

Export

The data included in this tile summarizes the items published in the journal in the JCR data year and in the previous two years. This three-year set of published items is used to provide descriptive analysis of the content and community of the journal. Learn more

Items

TOTAL CITABLE	% OF CITABLE OA
959	15.64%

Citations*

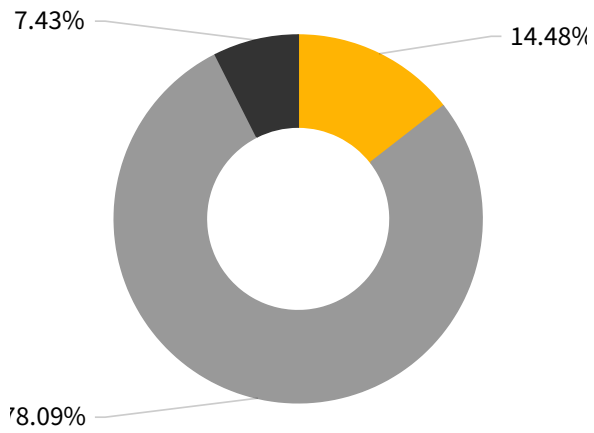
TOTAL CITABLE	% OF CITABLE OA
1,759	16.66%

15

CITABLE

GOLD OPEN ACCESS
150 / 14.48%

SUBSCRIPTION AND FREE TO READ
809 / 78.09%



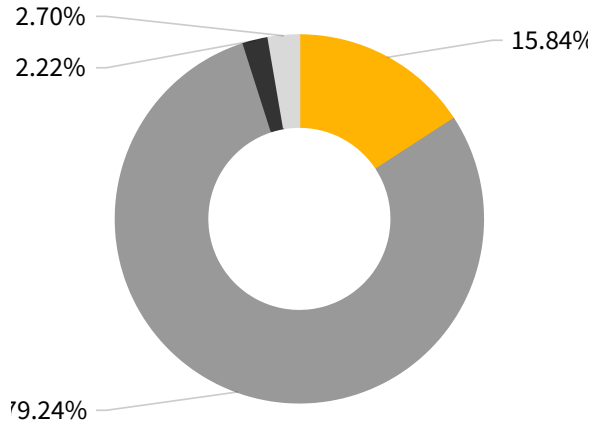
NON-CITABLE

OTHER (NON CITABLE ITEMS)
77 / 7.43%

CITABLE

GOLD OPEN ACCESS
293 / 15.84%

SUBSCRIPTION AND FREE TO READ
1,466 / 79.24%



NON-CITABLE

OTHER (NON CITABLE ITEMS)
41 / 2.22%

UNLINKED CITATIONS
50 / 2.70%

*Citations in 2024 to items published in [2022 - 2024]

Rank by Journal Impact Factor

Journals within a category are sorted in descending order by Journal Impact Factor (JIF) resulting in the Category Ranking below. A separate rank is shown for each category in which the journal is listed in JCR. Beginning in 2023, ranks are calculated by category.
Learn more

CATEGORY
PHYSICS, MULTIDISCIPLINARY

45/114

JCR YEAR	JIF RANK	JIF QUARTILE	JIF PERCENTILE
2024	45/114	Q2	61.0
2023	37/112	Q2	67.4

Rank by JIF before 2023 for PHYSICS, MULTIDISCIPLINARY

EDITION
Science Citation Index Expanded (SCIE)

JCR YEAR	JIF RANK	JIF QUARTILE	JIF PERCENTILE
2022	37/85	Q2	57.1
2021	39/86	Q2	55.23
2020	34/86	Q2	61.05
2019	45/85	Q3	47.65

Rank by Journal Citation Indicator (JCI)

15

Journals within a category are sorted in descending order by Journal Citation Indicator (JCI) resulting in the Category Ranking below. A separate rank is shown for each category in which the journal is listed in JCR. Data for the most recent year is presented at the top of the list, with other years shown in reverse chronological order. Learn more

CATEGORY

PHYSICS, MULTIDISCIPLINARY

38/114

JCR YEAR	JCI RANK	JCI QUARTILE	JCI PERCENTILE
2024	38/114	Q2	67.11
2023	36/112	Q2	68.30
2022	35/111	Q2	68.92
2021	35/110	Q2	68.64
2020	36/111	Q2	68.02
2019	46/109	Q2	58.26
2018	43/104	Q2	59.13

Citation network

Cited Half-life

5.5 years

The Cited Half-Life is the median age of the items in this journal that were cited in the JCR year. Half of a journal's cited items were published more recently than the cited half-life.

TOTAL NUMBER OF CITES

7,123

NON SELF-CITATIONS

6,704

SELF-CITATIONS

419

Citing Half-life

7.9 years

The Citing Half-Life is the median age of items in other publications cited by this journal in the JCR year.

TOTAL NUMBER OF CITES

18,966

NON SELF-CITATIONS

18,547

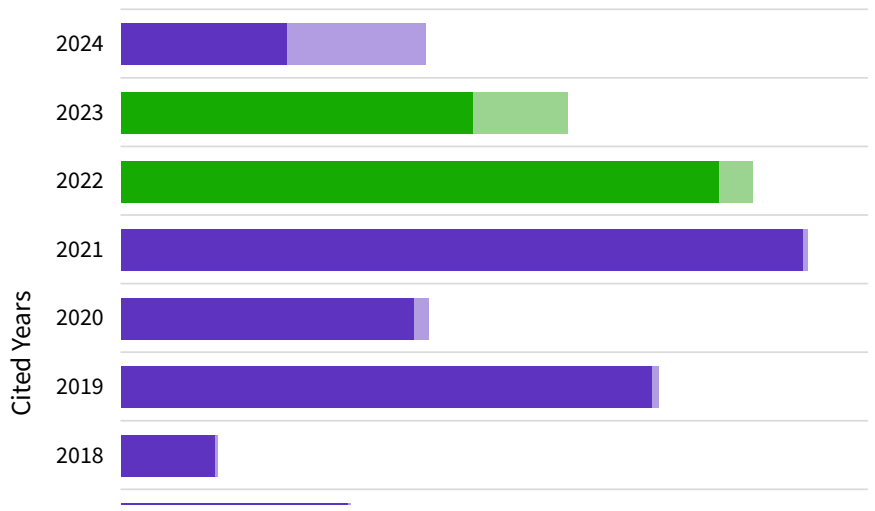
SELF-CITATIONS

419

Cited Half-life Data | Citing Half-life Data

Export

15

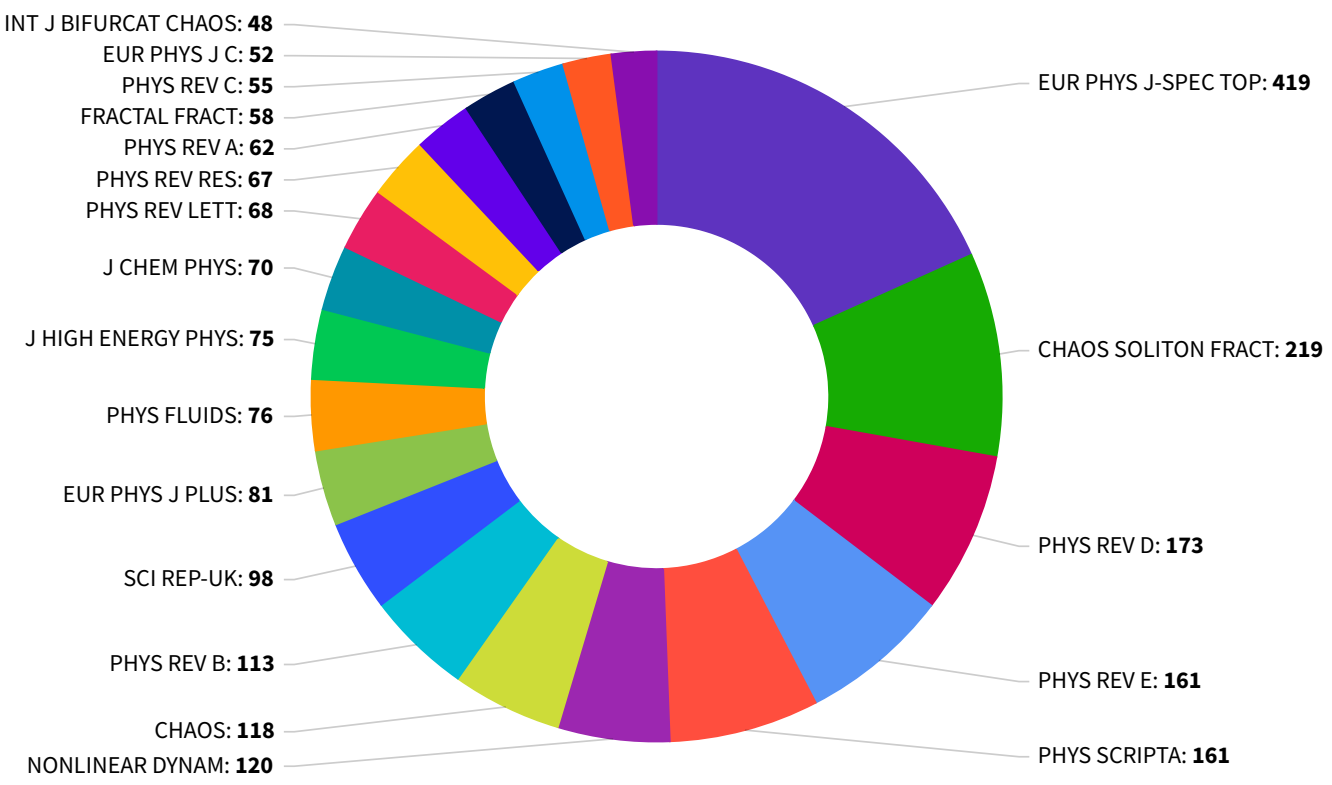


CITED YEAR	# OF CITES FROM 2024	CUMULATIVE %	# OF CITING SOURCES
All years	**7,123 citations**	**100.00%**	**1,623 sources**
2024	407 citations	5.71%	105 sources
2023	598 citations	14.11%	239 sources
2022	845 citations	25.97%	351 sources
2021	918 citations	38.86%	364 sources
2020	411 citations	44.63%	184 sources
2019	719 citations	54.72%	336 sources
2018	129 citations	56.54%	60 sources

Journal Citation Relationships

Cited Data Citing Data

Top 20 journals citing EUR PHYS J-SPEC TOP by number of citations



Content metrics

Source data

This tile shows the breakdown of document types published by the journal. Citable Items are Articles and Reviews. For the purposes of calculating JIF, a JCR year considers the publications of that journal in the two prior years. Learn more

Average JIF Percentile

Export

The Average Journal Impact Factor Percentile takes the sum of the JIF Percentile rank for each category under consideration 15 then calculates the average of those values. Learn more

344 total citable items

	ARTICLES	REVIEWS	COMBINED(C)	OTHER DOCUMENT TYPES(O)	PERCENTAG
NUMBER IN JCR YEAR 2024 (A)	277	67	344	21	94%
NUMBER OF REFERENCES (B)	11,154	7,496	18,650	316	98%
RATIO (B/A)	40.3	111.9	54.2	15.0	

ALL CATEGORIES AVERAGE
61.0

PHYSICS, MULTIDISCIPLINARY
61.0

Contributions by organizations

Export

Organizations that have contributed the most papers to the journal in the most recent three-year period. Learn more

RANK	ORGANIZATION	COUNT
1	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)	77
2	URAL FEDERAL UNIVERSITY	73
3	INDIAN INSTITUTE OF TECHNOLOGY SYSTEM (IIT SYSTEM)	50
4	RUSSIAN ACADEMY OF SCIENCES	40
5	VELLORE INSTITUTE OF TECHNOLOGY (VIT)	38
6	SARATOV STATE UNIVERSITY	34

Contributions by country/region

Export

Countries or Regions that have contributed the most papers to the journal in the most recent three-year period. Learn more

RANK	COUNTRY / REGION	COUNT
1	INDIA	290
2	CHINA MAINLAND	176
3	RUSSIA	173
4	GERMANY (FED REP GER)	104
5	FRANCE	89
6	USA	85
7	SAUDI ARABIA	52
8	TURKIYE	43
9	ENGLAND	42

Additional metrics

Eigenfactor Score

0.00631

The Eigenfactor Score is a reflection of the density of the network of citations around the journal using 5 years of cited content as cited by the Current Year. It considers both the number of citations and the source of those citations, so that highly cited sources will influence the network more than

Normalized Eigenfactor

1.41511

The Normalized Eigenfactor Score is the Eigenfactor score normalized, by rescaling the total number of journals in the JCR each year, so that the average journal has a score of 1. Journals can then be compared and influence measured by their score relative to 1. Learn more

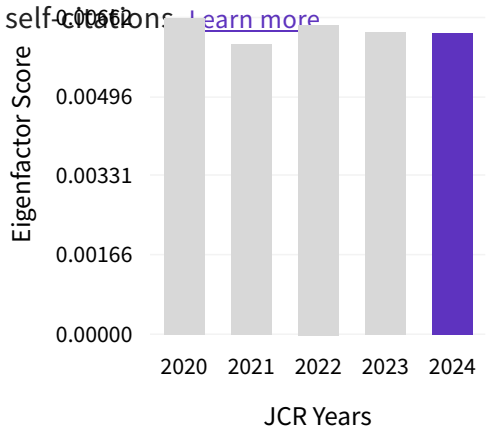
Article influence score

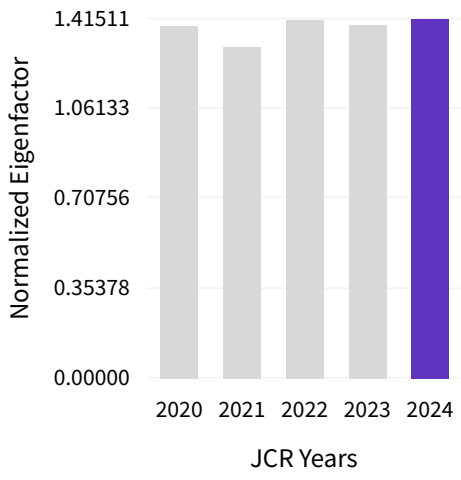
0.554

The Article Influence Score normalizes the Eigenfactor Score according to the cumulative size of the cited journal across the prior five years. The mean Article Influence Score for each article is 1.00. A score greater than 1.00

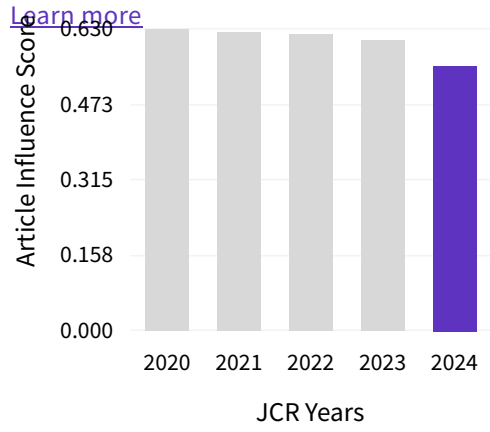
15

less cited sources. The Eigenfactor calculation does not include journal self-citations. Learn more





indicates that each article in the journal has above-average influence. Learn more

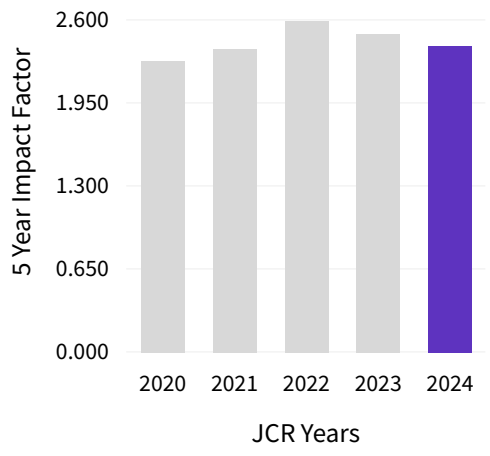


5 Year Impact Factor

2.4

View Calculation

The 5-year Impact Factor is the average number of times articles from the journal published in the past five years have been cited in the JCR year. It is calculated by dividing the number of citations in the JCR year by the total number of articles published in the five previous years. Learn more

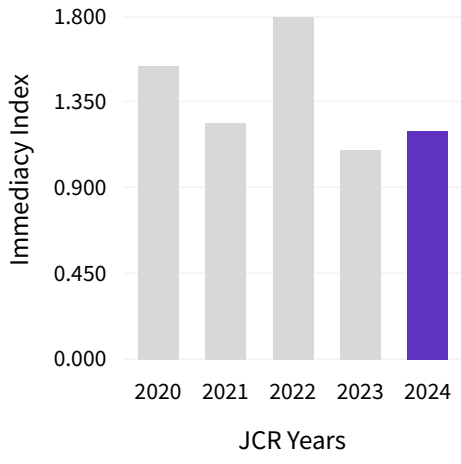


Immediacy Index

1.2

View Calculation

The Immediacy Index is the count of citations in the current year to the journal that reference content in this same year. Journals that have a consistently high Immediacy Index attract citations rapidly. Learn more



© 2025 Clarivate

Legal center

Privacy notice

Cookie policy

Terms of Use

Manage cookie preferences

Copyright notice

Accessibility

Help

15

SPRINGER NATURE Link

Log in

Menu Search Cart

Home > The European Physical Journal Special Topics > Editorial board



The European Physical Journal Special Topics

Publishing model: **Hybrid**

Kocaeli Üniversitesi Explore open access funding | Change institution

Journal menu

Editorial board

Managing Editors

Sandrine Karpe
EDP Sciences, Les Ulis, France

Vijala Kiruvanayagam
EDP Sciences, Les Ulis, France

Sabine Lehr
Springer-Verlag, Heidelberg, Germany

Editorial board

Shahriar Afkhami
Institute for Data Science Affiliated Faculty, New Jersey Institute of Technology, Newark, United States

G. Ambika
Indian Institute of Science Education and Research Thiruvananthapuram, Thiruvananthapuram, India

Balasubramanian Ananthanarayan
Centre for High Energy Physics, Indian Institute of Science Bangalore, Bengaluru, India

Kurt H. Becker
NYU Polytechnic School of Engineering, New York, United States

Klaus Blaum
Max Planck Institute for Nuclear Physics, Heidelberg, Germany

Roberta Citro
Dipartimento di Fisica "E. R. Caianiello, University of Salerno, Fisciano, Italy

Pere Colet
Institute for Cross-Disciplinary Physics and Complex Systems, Palma, Spain; Universitat de les Illes Balears, Palma, Spain

Miroslav Dramićanin
University of Belgrade, Belgrade, Serbia; Vinča Institute of Nuclear Sciences, Vinča, Serbia

Peter Hänggi
Fakultät für Physik, University of Augsburg, Augsburg, Germany

Baohua Ji
Institute of Biomechanics and Applications, Zhejiang University, Hangzhou, China

Professor Holger Kersten, Dr. rer. Nat.
Institut für Experimentelle und Angewandte Physik, Kiel University, Kiel, Germany

Sivarama Krishnan
Department of Physics, Indian Institute of Technology Madras, Chennai, India

Jürgen Kurths
Potsdam Institute for Climate Impact Research, Potsdam, Germany

Franck Lépine
Institut Lumière Matière, Villeurbanne, France

Grzegorz Litak
Department of Applied Mechanics, Lublin University of Technology, Lublin, Poland

Jingting Luo
Shenzhen Key Laboratory of Advanced Thin Films and Applications, Shenzhen University, Shenzhen, China

Norbert Marwan
Potsdam Institute for Climate Impact Research, Potsdam, Germany

Yves Pomeau
École Polytechnique, Palaiseau, France

Nuggehalli M. Ravindra

New Jersey Institute of Technology, Newark, United States

Günter Reiter
Physikalisches Institut, University of Fribourg, Fribourg, Switzerland

Helmut Schiessel
Center for Molecular and Cellular Bioengineering (CMCB), TU Dresden, Dresden, Germany

Filippos Sofos
University of Thessaly, Volos, Greece

Didier Sornette
Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

Eric Suraud
Laboratoire de Physique Quantique, Université Toulouse III - Paul Sabatier, Toulouse, France

Martin Weigel
Institute of Physics, Chemnitz University of Technology, Chemnitz, Germany

John G. Weisend II
European Spallation Source, Lund, Sweden

Eur. Phys. J. Spec. Top. (2025) 234:4829–4842
https://doi.org/10.1140/epjs/s11734-025-01510-5

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Prediction of pathological grade in prostate cancer: an ensemble deep learning-based whole slide image classification model

Gamze Korkmaz Erdem[1,a], Sevinc Ilhan Omurca[1,b], Esra Betul Cakir[2,c], and Busra Yaprak Bayrak[2,d]

[1] Faculty of Engineering, Department of Computer Engineering, Kocaeli University, 41300 Kocaeli, Turkey
[2] Faculty of Medicine, Department of Pathology, Kocaeli University, 41300 Kocaeli, Turkey

Received 16 December 2024 / Accepted 6 February 2025 / Published online 18 February 2025
© The Author(s) 2025, corrected publication 2025

Abstract Prostate cancer is a highly prevalent tumor among men and probably histologically represents one of the most common malignant tumors in the world. The automated analysis of tissues in histopathology has been advancing with higher accuracy rates using modern slide imaging techniques in combination with deep learning approaches. Today, pathologists can examine tissue samples in detail on high-resolution digital images thanks to whole slide imaging (WSI) images. This helps to accurately diagnose and evaluate the grades of prostate cancer. Tasks for analyzing WSI images with deep learning methods rely heavily on WSI images where tumor regions are marked by pathologists. This poses the challenge of marking WSI images for pathologists. In this study, we developed deep learning-based models with transfer learning support to classify prostate cancer grade in WSIs based on slide-level labels without the need for pathologists to mark the images. For this aim, we collected a real-life dataset which contains WSI image data of prostate cancer patients obtained from Kocaeli University, Department of Medical Pathology. The collected WSI data were labeled with prostate cancer grades only which are slide-level annotations, without the identification of regions of interest (ROIs) such as tumor regions within the tissue section. To demonstrate the contribution of the proposed, WSI-based classification models, they are compared with patch-based models regarding accuracy, precision, recall, and F1 score. Experimental results show that the proposed model provides high classification accuracy for prostate cancer detection over WSI images.

1 Introduction

Prostate cancer (PCa) remains one of the most common malignancies in men and is a leading cause of cancer-related mortality worldwide. Accurate diagnosis and staging are essential for determining the appropriate course of treatment, particularly in assessing the aggressiveness of the tumor. A key element in this process is the Gleason grading system, which classifies prostate tumors based on their histological architecture. The Gleason score provides valuable prognostic information and plays a crucial role in clinical decision-making, guiding treatment strategies such as active surveillance, surgery, or radiation therapy [1–3].

Despite its importance, manual evaluation of Gleason grades can be subjective, with significant variability between pathologists. This inconsistency can lead to inaccurate grading and, consequently, inappropriate treatment plans. These challenges have highlighted the need for more objective and reliable grading methods, particularly through the use of digital pathology and artificial intelligence (AI) [3, 4].

The introduction of whole slide imaging (WSI) has revolutionized pathology by enabling high-resolution digital scans of histopathological slides. This technology facilitates the application of deep learning models, especially convolutional neural networks (CNNs), to automate tumor classification and Gleason grading. AI-driven approaches have shown significant potential in reducing inter-pathologist variability and improving the accuracy of Gleason grading, offering a more consistent approach to prostate cancer evaluation [5, 6].

[a] e-mail: gamze.erdem@kocaeli.edu.tr (corresponding author)
[b] e-mail: silhan@kocaeli.edu.tr
[c] e-mail: esrabetul96@hotmail.com
[d] e-mail: busra.yaprakbayrak@kocaeli.edu.tr

However, using annotated WSI images presents its own set of challenges. For large tissue sections, manual tumor-region assessment is laborious and time-consuming, prone to inter-observer variability and uncertainty. Additionally, images of rare disease types may lead to imbalances between labeled positive and negative samples. Medical images can also vary greatly due to differences in devices, resolutions, and imaging protocols, complicating the process and affecting the generalization ability of AI models. Further, factors like noise and low contrast, as well as potential errors in manual annotations, can negatively impact the performance of classification methods [3–5].

Considering these challenges, it is important to utilize WSI images labeled only by tumor grade, without manually marking tumor regions [6]. This approach minimizes the errors associated with manual annotation and focuses on tumor grade labeling. To address these issues, we propose a novel approach using deep learning models, particularly leveraging transfer learning-assisted methods, for prostate cancer classification without the need for manual tumor marking. For this aim, we develop an ensemble deep learning-based WSI classification model to predict Gleason grade groups in prostate cancer. By combining multiple deep learning models, we aim to improve classification performance, enhance the model's capacity for generalization, and mitigate the impact of variables that degrade image quality, such as noise, low contrast, and differences in imaging techniques. Leveraging pre-trained CNN models and advanced machine learning techniques, our proposed approach offers a more efficient method for utilizing unlabeled WSI images for cancer classification and diagnosis. Ultimately, this method enhances the accuracy and reliability of Gleason grading from biopsy samples, supporting more informed clinical decision-making and improving patient outcomes.

To the best of our knowledge, no previous work has been done on using deep learning networks to classify cancer grade in prostate cancer images that are labeled at the slide level with Gleason grading only and do not include any markings for tumor areas.

Consequently, main contributions of the paper are as follows:

- Addressing the challenges of finding and working with marked pathology images, we present a novel approach for prostate cancer classification using transfer learning-assisted deep learning models without the need for manual marking of tumor regions.
- To increase classification performance and robustness, we employ ensemble learning techniques that involve combining numerous deep learning models. This method not only increases the model's capacity for generalization, but it also reduces the impact of variables that degrade image quality, such as noise, low contrast, and differences in imaging techniques. Our proposed approach offers an efficient method to use unlabeled WSIs for cancer classification and diagnosis.

The rest of the paper is organized as follows. In Sect. 2, the literature review of WSI image classification studies is presented. Section 3 briefs the methodology used in our study. In Sect. 3.3, the proposed model is described with the associated pipelines. In Sect. 4, the experimental design, the dataset, the evaluation metrics are described, and the experimental results are presented and discussed. Finally, the conclusions and the future directions of the study are given in Sect. 5.

2 Literature review

This section briefly introduces literature studies on WSI image classification tasks.

WSIs are large-size images that contain background and texture images. Additionally, it may contain various background noises. For this reason, it is often necessary to pre-process the images to train the models. Pre-processes applied to WSIs include data augmentation, noise removal with image filtering applications, and separation of WSI images into smaller images called patches. Looking at the literature, in most of the classification studies on medical images, the images were divided into patches and used to train the models [7].

For example, Mi et al. [8] suggested a two-stage architecture based on machine learning and deep learning techniques for the multi-class classification of digital pathological images of the breast. Region of interest (ROIs) are taken from WSIs and image cutting is applied to the ROIs in the first stage to obtain patches. Afterward, the InceptionV3 model was trained using the acquired patch images for patch-level categorization. To produce a heatmap for each WSI, the results of classifying each patch that is a part of that WSI were used. A final diagnosis with a diagnostic accuracy comparable to pathologists was possible because of the utilization of WSI-level heatmaps for feature extraction in XGBoost. BreakHis and BACH, two publicly available datasets, were used in testing the suggested method's robustness. Concerning patch-level and WSI-level classification, the respective overall accuracy ratings were 86.67% and 88.16%. In another study [9], Huang et al. present a deep learning architecture named BM-Net for the classification of breast cancer in digital pathology images. The architecture consists of a combination of MobileNet-V3 and a bilinear structure. BACH data set which has four classes of

breast tissue was used for experimental studies. In the study where images were divided into patches, the accuracy score obtained was 0.88. A deep convolutional neural network is used as a patch-level classifier by the authors in [8], while XGBoost is used as a WSI-level classifier. This approach allows for the automatic classification of breast digital pathology images into four classes: benign lesion, invasive carcinoma, ductal carcinoma in situ, and normal tissue. A heatmap including four predicted probability values was created by training InceptionV3 as a Patch-Level Classifier. The WSI-level classifier that they selected was XGBoost, based on a comparison of the classification accuracy of several classifiers.

Using one of the well-known pre-trained architectures is the simplest method for training CNNs with a small amount of input. Usually, pre-trained weights using ImageNet are used to initialize the models, and they are fine-tuned using images of pathology. One study that used this method was [10] where the authors fine-tuned the VGGNet, ResNet, and InceptionV4 models to determine the likelihood of tiny patches that were either TIL-positive or TIL-negative that were taken from WSIs of 23 different cancer types. They took larger superpatches and labeled them with three categories according to the ratio of TIL-positive area to improve the performance of region classification. To estimate the category, super-patches were split into eight squares, each of which was assigned a TIL-positive or TIL-negative classification. Following that, a correlation was noted between the pathologists' annotations and the CNN score (the total number of positive patches in the super-patch). The study achieved an accuracy score of 87.43% with the InceptionV4 model for all cancer types. In [11], training from scratch and transfer learning were compared. The three methods of training the VGG16 network that the authors compared were training the network from scratch, using transfer learning as a feature extractor, and fine-tuning breast cancer pictures to identify invasive ductal carcinoma. The model that was trained from scratch produced higher results in terms of accuracy (0.85), according to the findings. They were able to train a similar model (accuracy 0.81) ten times faster, though, thanks to transfer learning. Furthermore, compared to transfer learning by fine-tuning (accuracy 0.51), transfer learning via feature extraction (accuracy 0.81), which required retraining part of the convolutional blocks, produced better outcomes in less time. Researchers in [12] fine-tuned a pre-trained ResNet18 using the ImageNet dataset to predict 5-year overall survival in renal cell cancer. Each patch was given a probability score by CNN, and the class of a whole WSI was determined by averaging and classifying the scores of all related patches. They employed multivariable logistic regression analysis using CNN prediction in addition to single-stage classification with additional clinicopathological characteristics.

Remarkably, CNNs have shown themselves to be effective classifiers in the subject of histology. That being said, they may also be used with other machine learning (ML) classifiers. A CNN algorithm was created by the authors in [13] to automatically classify pathologic glioma (brain tumor) images into six classes. The CNN processes the images to produce patch-level output categories. Now, depending on the quantities and ratios of tumor kinds, the patch labels are run through a hierarchical decision tree to get the patient-level diagnosis. As a consequence, the outcome comprises the patient-level label and the image patch-label findings. Patch level accuracy is 86.5% and patient-level accuracy is 87.5% in the DenseNet-based research. The authors of [14] suggested MB-AI-His, a combination of deep learning and machine learning techniques, for the automated diagnosis and categorization of four pediatric MB subtypes. There are two stages to the diagnosing process. The first level divides each image into normal and pathological categories, and the second level (multi-classification level) divides the abnormal images that contain MB tumors into four subtypes of pediatric MB tumors. To extract important characteristics, three pre-trained deep CNNs (ResNet-50, DenseNet-201, and MobileNet) are used using transfer learning. Discrete wavelet transform (DWT) is used to extract time–frequency characteristics, which are then merged with these features. Lastly, multi-class classification is achieved by combining spatial information with five well-known classifiers: SVM, k-NN, Ensemble Subspace Discriminant, and Linear Discriminant Analysis. In [15], the authors present a similar strategy. In comparison, binary classification is significantly simpler than multi-class classification of the four types of childhood MB. Very few studies have looked at this multi-class classification challenge. Their process comprises feature fusion and reduction using the DWT approach, spatial DL feature extraction using ten optimized CNN architectures, and feature selection after that. A bidirectional Long-Short-Term Memory classifier is used for classification. According to the authors of [16], they were the first to suggest a technique based on CNNs for identifying pancreatic ductal cancer in WSIs. They used a heatmap of the chance of malignancy in conjunction with anticipated patches and InceptionV3 as a patch-level classifier. At this phase, a Light Gradient Boosting Machine (LightGBM) was trained for slide-level classification using statistical characteristics that were taken from WSI heatmaps.

Studies in the literature are mostly on multi-class classification. A multi-class classifier based on deep learning was developed in [17] to identify more types of lung abnormalities, including lung cancer. They also included organizing pneumonia and pulmonary tuberculosis, which frequently require surgical inspection to be identified as malignancy. They used EfficientNet and ResNet to do patch-level categorization. To combine patch predictions into a slide-level classification, mean pooling and majority voting were compared. Additionally, a two-stage aggregation method was used to rank the cancer tissues on slides. A deep-learning classifier was built and used by researchers in [18] to categorize lung carcinoma samples into nine tissue groups. Then calculated spatial features from the retrieved characteristics that characterize the tumor microenvironment's makeup, and then combined these features with clinical data to predict tumor mutation and patient survival. To categorize breast cancer into four categories,

the authors in [19] handled the challenge of binary classification into benign and malignant lesions, with further separation into eight subtypes using modified EfficientNetV2 architecture using the BreakHis dataset. The scientists used the Grad-CAM approach to show that CNN models paid attention to the cancer cell organization patterns while discriminating histopathological tumor pictures of different grades since neural networks act like black-box models. Grad-CAM was also used in [20], where the authors presented the model's findings to pathologists using this approach to give interpretability and an approximative visual diagnosis. The algorithm classified H&E-stained tissue patches into five categories of liver lesions, cirrhosis, and almost normal tissue using three neural networks that were fine-tuned using a custom dataset.

Transformer-based methods are also seen in the literature. In one of these studies [21], a novel self-supervised learning (SSL) framework called Semantically Relevant Contrastive Learning (SRCL) was introduced for feature extraction from histopathological images. The proposed hybrid model combines CNNs and Swin Transformer architectures to capture both local and global features effectively. SRCL enhances traditional contrastive learning by selecting semantically relevant positive pairs from a large memory bank, improving feature diversity and representation. The model was pretrained on over 15 million patches from the TCGA and PAIP datasets. In the study, tasks such as patch retrieval, patch classification, weakly supervised WSI classification, and mitosis detection were discussed. The F1 score obtained on the patch classification is 0.8988 and an accuracy score of 0.9266. For WSI classification, an accuracy of 0.947 and an AUC score of 0.988 were achieved with the TCGA-RCC data set.

According to studies in the literature, hybrid models are also recommended. In [22], YOLOv5 models were used for tumor location detection, and ResNet50 models were used for feature extraction. Model training and validation were performed using the Cancer Genome Atlas (TCGA) dataset. It was worked with patches of two different sizes: 256×256 and 512×512. In the study where various data augmentation techniques were used, an accuracy score of 97.2% was achieved.

These days, it is increasingly common to apply attention methods to CNNs to improve performance. The Divide-and-Attention Network (DANet), for the categorization and grading of colorectal and breast malignancies, was suggested in [23]. The initial diseased picture, the image of the nuclei, and the image of the non-nuclei are the three inputs to this network. A segmentation model of nuclei is used to acquire the pictures of nuclei and non-nuclei. A Nuclei-Guided Network (NGNet) for the grading of breast invasive ductal carcinoma was developed by the authors of [24] using a methodology close to DANet. In contrast to DANet, NGNet requires just two input images: the initial image and the segmentation-obtained nucleus image.

The extensive literature analysis also shows that the use of datasets labeled with tumorous areas is common in the classification of WSI images. Studies using only the limited information of whether the WSI image is tumorous or not without any marking of the tumorous area in prostate cancer classification are very rare. However, in real life, there is a cost for pathologists to mark the tumorous area on WSI images. Our proposed model aims to fill an important gap in the literature as an end-to-end automated solution in this context.

3 Methods

In this section, a comprehensive overview of deep learning models that have been used is presented.

3.1 Convolutional neural networks

CNN models have been used extensively in the literature in image-based medical diagnoses and predictions due to their proven effectiveness in image recognition tasks [25]. In CNNs, an input image is first transformed into a feature map, which is then processed through multiple convolution and pooling layers to produce a prediction label. In convolution operations, a set of filters is employed to the input image, in this way, a feature map that highlights fundamental aspects of the input image is extracted. Then the pooling layer downsamples the feature map to reduce its size while keeping the most informative features. The extracted feature map passes through multiple additional convolutional and pooling layers to capture complex and hierarchical relationships between different features in the input data [25].

3.2 Transfer learning and pre-trained models

Training deep forward neural network architectures for image classification tasks requires continuous updating of a lot of parameters, a large amount of training data, and a long Graphics Processing Unit (GPU) run time. This leads to high computing resource costs and low efficiency. Transfer learning is a machine learning method that utilizes a pre-trained model which is created for a specific task to solve different tasks. Transfer learning stands out with its main advantages among several machine learning techniques: the ability to work with small amounts of local training data, low requirement of computational resources, and time-saving [26].

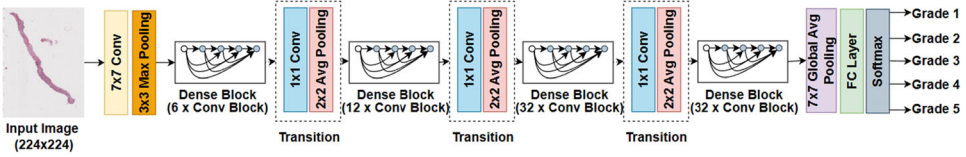


Fig. 1 The architecture of DenseNet-169 used to implement the proposed method

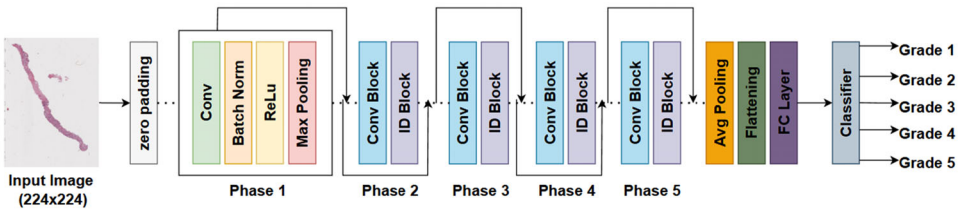


Fig. 2 The architecture of ResNet50v2 used to implement the proposed method

In this study, transfer learning was applied by fine-tuning several CNN models such as DenseNet169, DenseNet121, and ResNet50V2. In this way, the pre-trained weights and architectures of the models were utilized, resulting in significant reductions in training time while improving prostate cancer prediction accuracy.

DenseNet is a special convolutional neural network architecture that uses dense connections from each layer to all other layers. DenseNet169 is one of the DenseNet variants that has recently shown impressive results in image classification tasks. DenseNet169 consists of 169 low-parameter layers, transition layers, and finally a classification layer. Similarly, DenseNet121 consists of 121 convolution layers with each convolution block followed by dense blocks and transition layers [27]. The architecture of DenseNet169 model which are implemented in our study are shown in Fig. 1. The architecture of DenseNet121 and DenseNet169 differ in the number of convolutional layers in the third and fourth Dense Blocks. In the third Dense Block, there are 24 convolutional layers for DenseNet121 and 32 convolutional layers for DenseNet169. In the fourth Dense Block, there are 16 and 32 convolutional layers for DenseNet121 and DenseNet169, respectively.

ResNet, like Densenet, is a robust convolutional neural network architecture that is widely used in medical image classification as well as general-purpose image classification. In this paper, a pre-trained ResNet50v2 [28] which employs 50 convolution layers, 3×3 kernels, bottleneck layers, and multiple residual connections was applied to predict the grade of prostate cancer from the pathology images. As the last grade, the fully connected layers generate the probabilities of prostate grades. The architecture of ResNet50v2 is reflected in Fig. 2.

3.3 Proposed model

We propose an annotation-free ensemble deep learning-based WSI classification model. Our proposed model requires only slide-level annotations such as the Gleason grade of the prostate cancer. For this aim, patch-based classification was performed first. Instead of using all WSIs directly, the purpose of patch-based classification is to divide the WSIs into small pieces called 'patches' and generate class predictions for these patches. The models trained with small patches are used as feature extractors in WSI-based classification.

For patch-based classification, which is shown in Fig. 3, we first extracted 224×224 patches from the original WSIs. Each patch was labeled with the label of the WSI to which it belongs. The patches were then given as input to the pre-trained DenseNet121, DenseNet169 and ResNet50V2 models. Each of the trained models was saved separately for use in WSI-based classification. Finally, predicted grade labels for each patch were achieved.

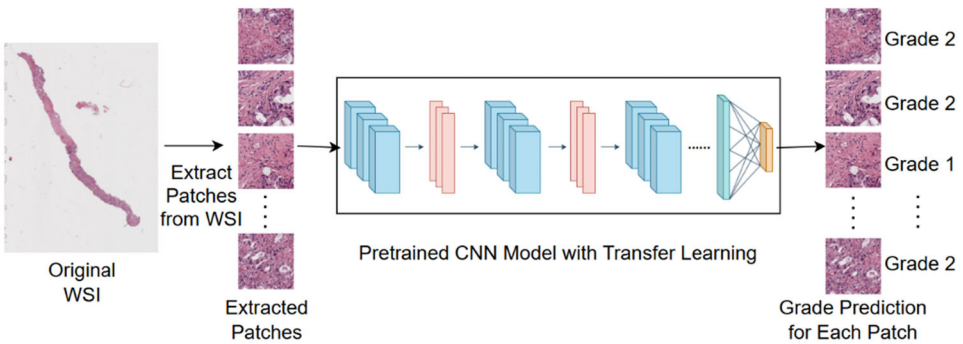


Fig. 3 The patch-based classification pipeline

After the patch-based classification was completed, WSI-based classification was performed. In WSI-based classification, models trained and saved using patch images are used to extract features from individual patches. The goal is to aggregate the patch-based predictions for each WSI and determine the final class prediction of the WSI by majority voting and weighted majority voting.

The prediction of patch classes differs between majority voting and weighted majority voting techniques. In order to use a majority vote, the most expected class is chosen. Every model in this method has the same weight. The weight values allocated to the models are employed in weighted majority voting to predict patch classes. Considering each model's performance independently, the weights are established. The greatest weighted total prediction is calculated by multiplying these values by the prediction values and then adding them together.

To use majority voting for WSI-based classification, all patch images were collected into files called after the WSI that they belonged to. Then, for label prediction, each patch image was fed into pre-trained models. All of the patches that belonged to each WSI's prediction label were saved independently. WSI's final label is the most predicted class label for patches extracted from the WSI. WSI-based classification and majority voting pipeline is given in Fig. 4.

As a second approach, we aimed to evaluate each patch of a given WSI using different CNNs to provide diversity. The decisions of different CNN base classifiers for each patch of a given WSI are combined with the cooperative learning hypothesis to form a stronger classification hypothesis. Thus, independent classifiers are created that evaluate each patch as if they were separate expert pathologists, and the decisions of these classifiers are voted on. For this aim, every model was first taken into consideration independently in WSI-based classification. Next, all the models combined were used to apply weighted majority voting. The patches taken out of each WSI were used as input for each model in this way. The predicted labels for the patches are each model's output. Following that, weighted majority voting was used using these labels. By summing the expected labels for each extracted patch by the model weights and finding the largest weighted sum, the final grade label of each WSI is obtained. The pipeline of WSI-based Classification with Ensemble Learning is reflected in Fig. 5.

The weights of the models were assigned to the ensemble decision according to the classification errors obtained when they were applied as patch-based classifiers.

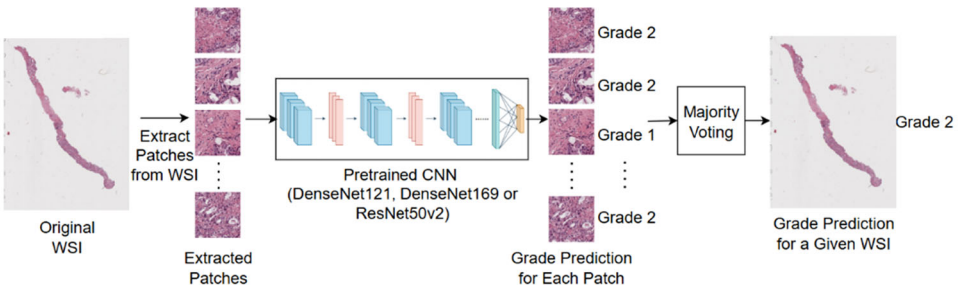


Fig. 4 The WSI-based classification and majority voting pipeline

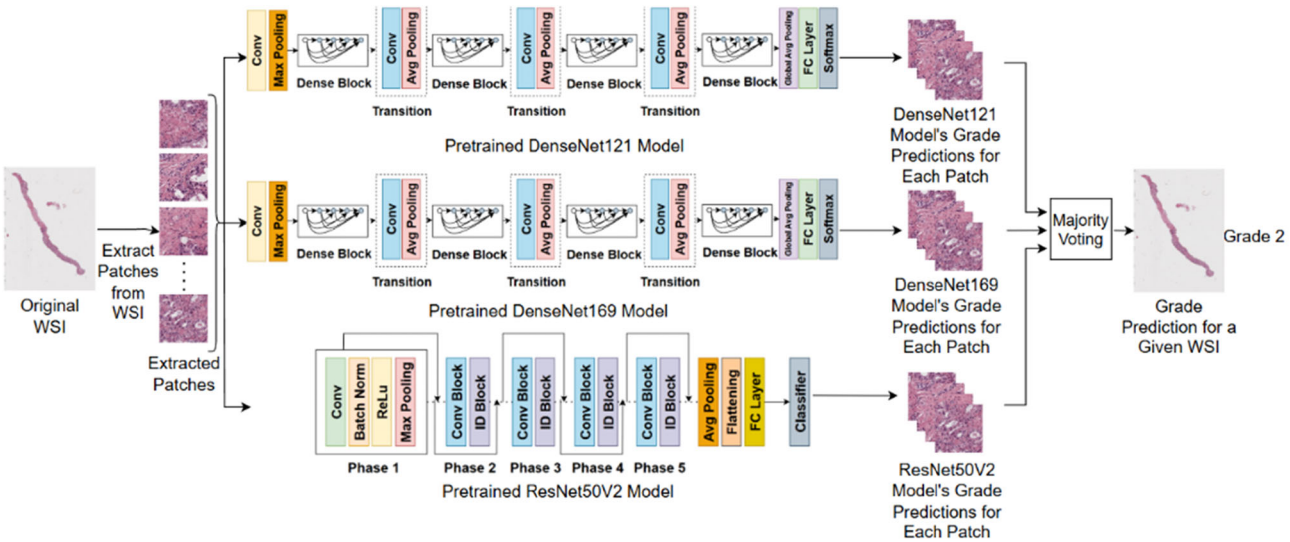


Fig. 5 WSI-based classification with ensemble learning pipeline

4 Results and discussion

4.1 Dataset and preprocessing

The prostate cancer tissue images used for the study were obtained from the Medical Pathology Department of Kocaeli University, Faculty of Medicine. The WSI images are in RGB.tif format, and each image is approximately 10,000 × 10,000 pixels in size. They were labeled with 5 different class labels as Grade 1, 2, 3, 4, and 5 by expert pathologists. All of these classes belong to one type of prostate cancer and each class represents a cancer grade. As the grade level increases, the cancer type worsens. The total number of WSIs used in the study is 36, 34, 33, 40 and 38 for each class, respectively, due to the loss of most of the texture image due to black spots or blurring.

The same dataset was used for both WSI-based and patch-based classification, but the file organization of the data differed for the two classification methods. For patch-based classification, each patch image was stored in files with the name of the class assigned to it. For WSI-based classification, a file named WSI was kept under each class file, and patches extracted from the corresponding WSI were kept under this file. This also preserved the WSI name information to which the patches belonged.

The original images cannot be directly processed by deep learning models due to their large size. In addition, there are parts of the original images that do not contain texture information, black areas, or regions containing noise. The preprocessing steps to make the data suitable for the models with minimal data loss are as follows.

- *Cleaning image backgrounds*. First, the images were converted from.tif format to.png format and then from RGB to gray format for easier processing (Fig. 6b). Median Blur was used to reduce noise in gray images (Fig. 6c). Otsu algorithm [7, 8] was applied to the blurred images. The Otsu algorithm selects an optimal threshold by analyzing the image histogram. The image histogram represents the distribution of intensity values of pixels. The algorithm calculates the probability of pixels belonging to two different classes and determines a threshold value that minimizes the variance between the classes. The image is segmented using the optimal threshold. The pixels are divided into two classes according to the optimal threshold value and thus objects or regions in the image are distinguished. The application of the Otsu algorithm results in binary mask images. In the mask images, the white areas represent texture regions (Fig. 6d).
- *Contour detection from images*. Mask images are used to determine the boundaries of tissue regions in images and to extract the regions of interest by surrounding them with a suitable area (Fig. 7a). The OpenCV library was used to outline the texture regions. The resulting point list is used to draw contour points on the original image. Most of the images used in this study contain fragmented texture regions. Therefore, all contour regions found for each image were recorded and processed separately (Fig. 7b). By detecting the contour points, it will be possible to determine and record rectangular areas suitable for these points. For each contour area, the coordinates of the upper, lower, left, and right endpoints were determined, thus creating the boundaries of the rectangular areas to be cut. All images cut according to the boundaries were saved (Fig. 7c).
- *Extracting patches from WSIs*. The large size of the data used in the study has led to the need to divide the images into small, fixed-size parts. This process provides advantages such as providing standard input sizes for training deep learning models, optimizing memory usage, reducing model training time, and focusing more on image details. From the sub-images extracted according to the contour regions, 224 × 224 pixel-sized fragments were cropped (Fig. 7d). The reason why the size is 224 × 224 pixels is to match the input size of the models to be trained. For the cropping process, it can be imagined that a window of the relevant size is scrolled step by step over the image, and the corresponding parts are cropped and saved. There is no overlap between the cropped parts. The proportion of texture region in the resulting fragments is not checked during the cropping process. After the process, all the fragments obtained are examined and the fragments containing less than 60% of the texture image are eliminated.
- *Application of data augmentation techniques*. Data augmentation increases the generalization ability of the model by adding diversity to the existing data set. This is especially important when the dataset is small. The data augmentation techniques applied to the training dataset in this study are rotation, brightness modification, x- and y-axes shifting, horizontal and vertical rotation, shearing, zooming, and rescaling (Fig. 7e). Only rescaling was applied to the validation and test datasets.

After preprocessing, the resulting data set is unbalanced. The most important reason for this is that there is more data belonging to the grade 5 class and less data belonging to the grade 1 class. To overcome this imbalance, the number of data belonging to all classes was organized according to the grade 1 class, which has the least data. As a result, a total of 17,000 patch images were obtained. The labels of the patch images were determined as the labels of the WSI to which they belonged. The dataset is divided into training, validation and test sets with 80%, 10% and 10%, respectively. While the Patch-Based Classification and WSI-based classification performances were defined according to the validation set; the classification results reflected in the confusion matrix were defined according to the test set.

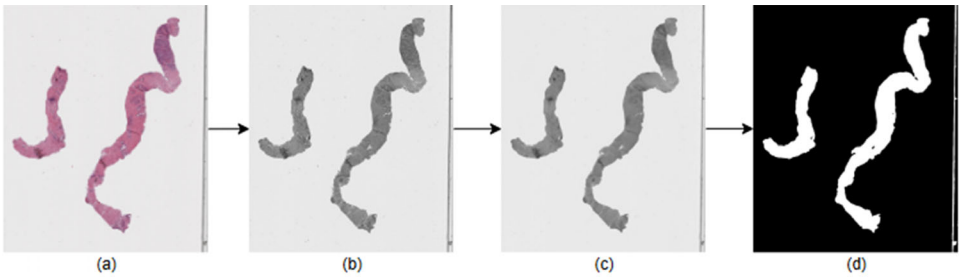


Fig. 6 Cleaning original image background and getting mask image: **a** original image, **b** gray-scale image, **c** image with median blur filter applied, **d** mask image obtained by applying Otsu algorithm

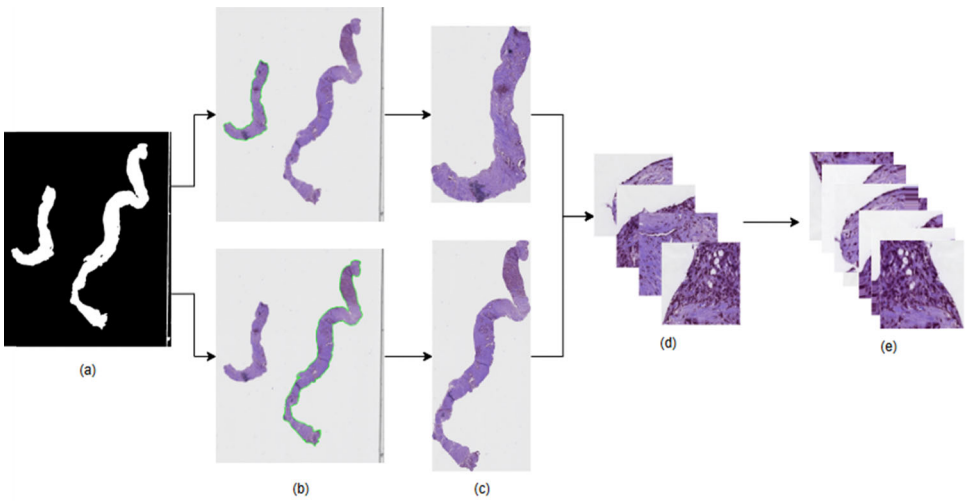


Fig. 7 Obtaining patches and data augmentation using the mask image and original image: **a** mask image, **b** contour detection from images. Contours are shown in green. **c** Cutting tumor regions according to contour areas, **d** extracting patches from WSI, **e** data augmentation

4.2 Experimental design

DenseNet169, DenseNet121, and ResNet50V2 models are pre-trained with the ImageNet dataset. Since the dataset obtained in this study contains medical images and is quite small compared to the ImageNet dataset, the models were trained using transfer learning. The experiments were implemented in Python using Google Colaboratory Pro and Keras libraries with Tesla T4 GPU.

In the study using deep learning-based models, both patch-based and WSI-based training were performed. For this purpose, patch images obtained as a result of preprocessing were first used, thus patch-based classification was performed. The input size of the models used for patch-based classification was set to 224 × 224. The models were trained in two stages. In the first stage, all layers except the last layer of the models were frozen and the last layer was removed. A convolutional, a global average pooling, and a softmax layer were added instead of the removed layer, respectively. In the convolutional layer, the activation function is ReLu and the number of neurons is 256. The learning rate value used at this stage is 0.001. If there is no improvement in the validation loss value of the model for 3 consecutive epochs during training, the current learning rate value is multiplied by 0.25 to create a new learning rate value (Model Checkpoint), and the value is gradually reduced. In addition, the model with the highest accuracy during training was saved, and if the validation loss value did not improve by at least 0.001 for 15 consecutive epochs, the training was stopped (Early Stopping). The models were trained for 50 epochs at this stage.

In the second stage, the model was retrained by unfreezing the frozen layers. In this stage, the learning rate value was set to 1e-5 and remained unchanged throughout the training. The 'ModelCheckpoint' and 'EarlyStopping' features were used in this stage with the same settings. The models were trained for 20–40 epochs at this stage.

The models were trained using the Adam optimization algorithm in both stages. In the algorithm, momentum was set to 0.9; the batch size was 16 and the loss function was 'categorical_crossentropy'.

4.3 Evaluation metrics

The obtained results are evaluated considering Accuracy, Precision, Recall, and F1 score metrics. Confusion matrices of the models are also provided to discuss the classification results in more detail.

The components of the confusion matrix are: True positive (TP) indicates properly predicted a particular cancer grade. True negative (TN) represents accurately predicted remaining cancer cases. FN indicates a particular

Table 1 Performance metrics for models based on patch-based and WSI-based classification

Metrics	Type of classification					
	Patch-based classification			WSI-based classification		
	DenseNet169	DenseNet121	ResNet50V2	DenseNet169	DenseNet121	ResNet50V2
Accuracy	77.24	67.35	75.76	91.16	82.87	87.85
Precision	77.26	68.14	76.00	92.19	85.23	88.75
Recall	77.24	67.35	75.76	91.16	82.87	87.85
F1 Score	77.18	67.24	75.61	90.59	82.84	87.25

cancer grade wrongly predicted as the remaining cancer grades. FP indicates a particular cancer grade incorrectly predicted as the remaining grades by the classification model. The formulas of the metrics utilized to evaluate the performance of the models are given as follows.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$F1\text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Precision gives the accuracy of positive predictions, while recall measures the model's ability to capture all relevant instances. Accuracy which is defined as the ratio of correctly predicted instances to the total dataset is an overall performance metric. *F1* score, calculated as the harmonic mean of both precision and recall, is a standard metric for classification tasks.

Accuracy, precision, recall and *F1* score metrics have been used because they can offer a complete perspective on the performance evaluation of classification tasks in machine learning. However, the *F1* score is important in medical image classification as it provides a balanced measure of performance, particularly in scenarios involving imbalanced datasets and critical errors. Unlike accuracy, which can be misleading in class-imbalanced settings, the *F1* score combines precision and recall into a single metric, ensuring a more comprehensive evaluation.

4.4 Experimental results

The performance scores obtained in the patch-based and WSI-based classification models are presented in Table 1.

Considering the results shown in Table 1, it is seen that WSI-based models are more successful than patch-based models regarding all of the performance metrics. In patch-based classification models, the model with the highest accuracy was DenseNet169 with an *F1* score of 77.18, while the least successful model was DenseNet121 with an *F1* score of 67.24. The fact that DenseNet169 has more layers and parameters compared to DenseNet121 allows it to achieve higher accuracy rates. According to the results reflected in the table, it is seen that the ResNet50V2 model has success rates close to DenseNet169. Although the ResNet50V2 model has more parameters, its layer structure is different. It has more sparse connections due to residual connections. Therefore, ResNet50V2 was slightly less successful than DenseNet169.

A similar ranking among the models is also observed in WSI-based methods. However, the success rates increased significantly compared to patch-based methods. The proposed WSI-based classification method improved the prediction performances by 13.41% for DenseNet169, 15.5% for DenseNet121 and 11.64% for ResNet50V2, respectively. In the WSI-based classification models, the majority voting strategy for patches significantly maintains the overall classification accuracy of the WSI image, even if some of the WSI patches are mislabeled, and reduces the negative impact of mislabeled patches. This improves the prediction performance of the proposed model compared to patch-based models.

In WSI-based classification, each patch belonging to a distinct WSI image could be the input of a single CNN network. The predictions generated by the CNNs for the patches were combined to predict the WSI label. After this grade, we aimed to boost classification performance of the WSI-based classifier using an ensemble learning approach. We utilize DenseNet169, DenseNet121 and ResNet50V2 as base learners to provide diversity.

Table 2 Performance metrics based on Ensemble learning

Metrics	Weighted Majority Voting	Majority Voting
Accuracy	92.82	88.95
Precision	93.69	90.89
Recall	92.82	88.95
F1 score	92.55	88.75

For WSI-based classification, each model was first evaluated separately. For each patch, the predicted class labels from the models are used to generate predicted class labels for the WSI to which the patches belong. Then, the effect of using all models together on the classification success was examined. At this grade, a patch image was given to all models and estimated class labels were obtained from each model separately. These predictions were then combined to obtain the final class predictions for the WSIs to which the patches belonged. At the grade where all models were used, both majority voting and weighted majority voting methods were used to measure the classification success. The success rates obtained using all models according to the voting methods are given in Table 2.

The combination of the three models in majority voting resulted in a score close to the average of the accuracy scores obtained when the models were used separately. Furthermore, weighted majority voting was found to be more successful than majority voting. This is because more weight is given to the models that are more successful than the patch-based classification success and have a greater impact on the voting result.

The resulting confusion matrices are provided for a better understanding of the model performances of the test set. Figure 8 displays the confusion matrices of patch-based classification models.

Looking at the confusion matrices of patch-based classification, it can be seen that DenseNet169 is the most successful model. It gives the most accurate results, especially for Grade 1, Grade 3, and Grade 4. In Grade 2, the ResNet50V2 model outperforms the other models in terms of classification rate, whereas in Grade 5, the DenseNet121 model performs better. Grade 4 had the lowest classification of any model when it came to class-based classification. This is because Grade 4 data are interpreted as being comparable to Grade 3 and Grade 5, and the models are unable to distinguish between these classes at the appropriate level. During categorization, the models misinterpreted the data for the other grades as data from other grades. As a result, patch-based classification has a lower success rate.

The confusion matrices based on the test set of three models for WSI-based classification are given in Fig. 9.

Based on the computed confusion matrices for WSI-based classification, DenseNet169 showed up as the most effective model. Considering classification based on classes, it has been observed that all three models accurately predict the grades 1, 2, 3, and 4. However, there are some differences between models for the Grade 5 prediction. Compared to the other models, the ResNet50V2 model shows higher confusion in this class. The success rate for the Grade 4 class in patch-based classification is observed increasing, although Grade 5 can be classified with reduced accuracy in all models. Grade 5 data is especially confused with Grade 1 and Grade 3 data for all models. Although the performance of the models is better than patch-based classification, correcting this situation in Grade 5 will contribute to improving model performance.

The confusion matrices based on the test set obtained by combining three models according to the voting method are given in Fig. 10.

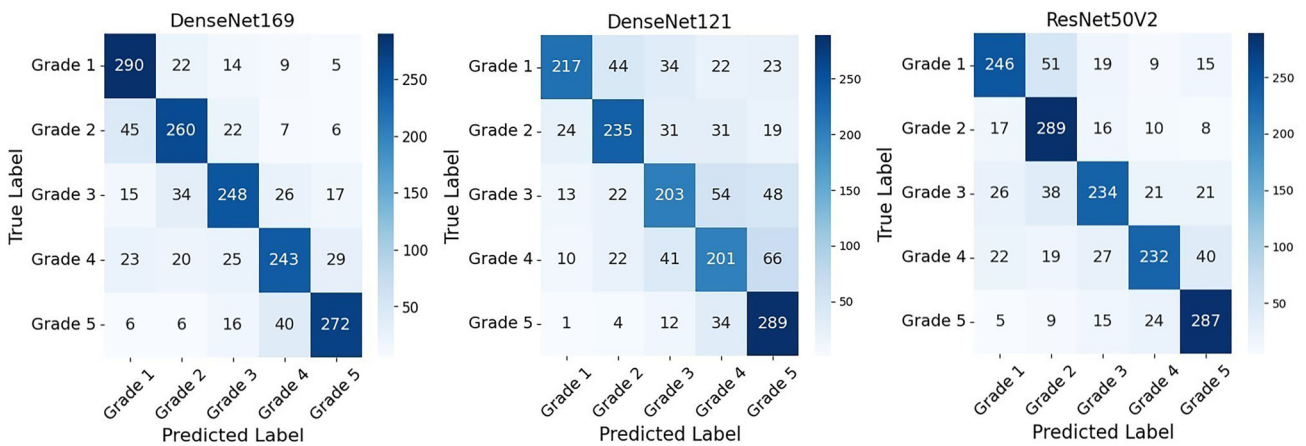


Fig. 8 Confusion matrices by models for patch-based classification

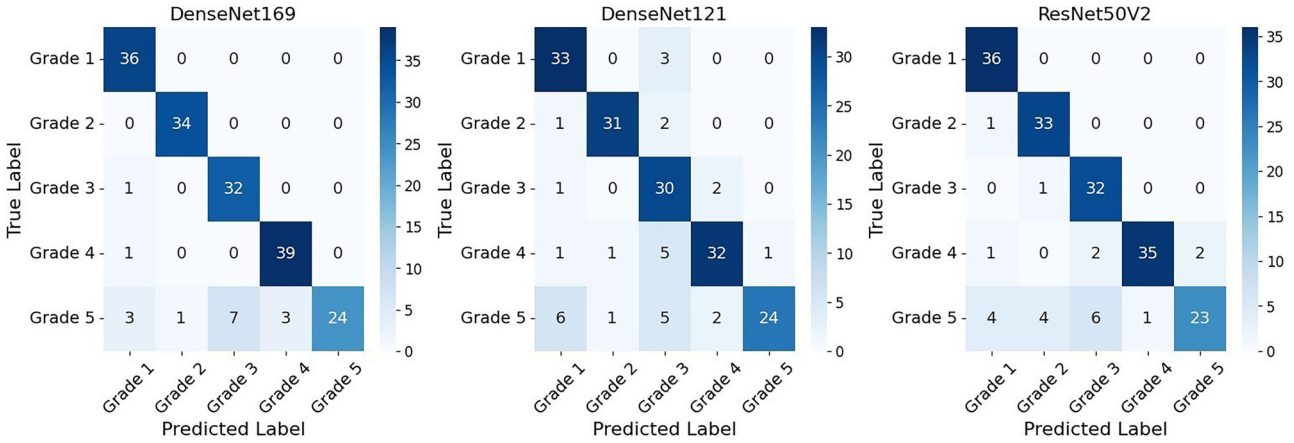
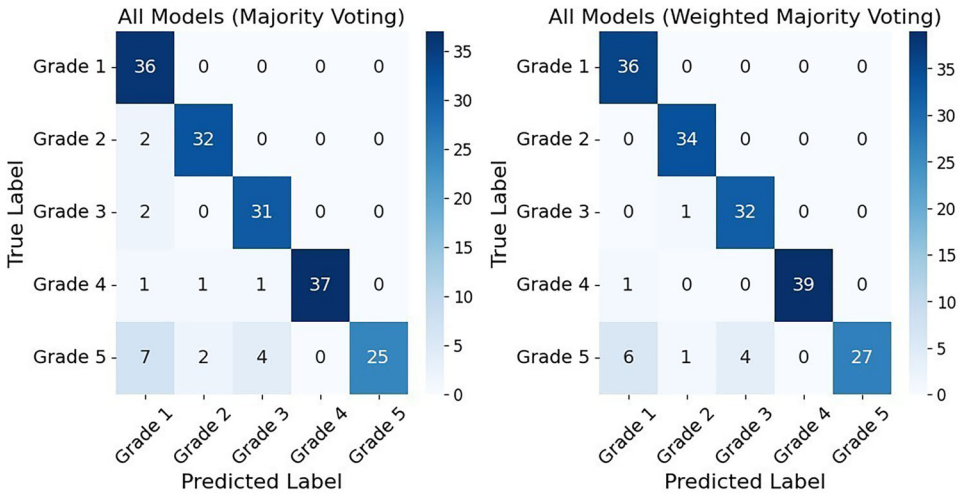


Fig. 9 Confusion matrices by models for WSI-based classification

Fig. 10 Confusion matrices according to voting methods



The use of ensemble methods according to the confusion matrices obtained with the voting methods has been an effective strategy to improve model performance. In particular, it is observed that the Weighted Majority Voting method gives better results in Grade 2 and Grade 5 classes compared to WSI-based classification, and both voting methods reduce misclassifications for Grade 5 class. Weighted Majority Voting is more successful than Majority Voting for all classes. The main reason for this is that more weight values are assigned to the more successful model, thus increasing the impact of the more successful model on the overall classification. This strategy can be effective in reducing misclassifications, especially in difficult classes (e.g. Grade 5), by enabling effective use of model combinations.

5 Conclusion

In this study, we proposed a novel deep learning-based approach for classifying prostate cancer grades directly from Whole Slide Imaging (WSI) images, without the need for manual annotation of tumor regions by pathologists. By leveraging transfer learning, ensemble learning approaches, our model achieved high classification accuracy for prostate cancer detection based on slide-level labels on a real-life dataset obtained from Kocaeli University, medical pathology department. Our study, conducted under the supervision of expert pathologists, simplified the workflow for pathologists and reduced the burden of identifying regions of interest (ROIs) in WSI images. There was no labeling on any of the images used for this study. The patches obtained for each WSI were assigned the label of the WSI they belong to. The results demonstrate that even in the lack of labeling, WSIs can be classified with good accuracy. The results demonstrate that the WSI-based classification model outperforms the patch-based model in

terms of accuracy, precision, recall, and *F1* score, highlighting its potential to enhance diagnostic workflows in prostate cancer diagnosis.

This study demonstrates significant scientific prospects due to its relevance, innovation, and practical applications in improving prostate cancer diagnosis. By utilizing slide-level labels instead of requiring manual and detailed region-of-interest annotations, the proposed approach significantly simplifies the annotation process for pathologists. This innovation addresses a major bottleneck in medical data preparation, making the methodology more efficient and scalable. Through the application of deep learning techniques, the study offers a promising solution to determine prostate cancer grading and enhance diagnostic workflows.

It is predicted that the success of the models will increase with the increase in the amount of data. Therefore, future work is going to focus on expanding the dataset, investigating the effectiveness of transformer architectures and the Yolo family algorithms instead of CNN architectures, and developing systems that provide higher accuracy rates by incorporating additional clinical metadata into the system.

Acknowledgements The authors would like to thank to the Department of Pathology at the Faculty of Medicine, Kocaeli University, for their invaluable support and contributions to this study. Their expertise and guidance have been instrumental in the completion of this work.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The data used in this study are not open for sharing due to data privacy.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval The study protocol was approved by the Ethical Committee of Non-invasive Clinical Research of Kocaeli University (Decision No: GOKAEK-2025). This study was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all participants, and no procedures involving intervention or invasion of patient privacy were performed.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. B. Liu, Y. Wang, P. Weitz, J. Lindberg, J. Hartman, W. Wang, L. Egevad, H. Grönberg, M. Eklund, M. Rantalainen, Using deep learning to detect patients at risk for prostate cancer despite benign biopsies. iScience **25**, 104663 (2022). https://doi.org/10.1016/j.isci.2022.104663
2. K. Ozbozduman, I. Loc, S. Durmaz, D. Atasoy, M. Kilic, H. Yildirim, T. Esen, M. Vural, M.B. Unlu, Machine learning prediction of Gleason grade group upgrade between in-bore biopsy and radical prostatectomy pathology. Sci. Rep. (2024). https://doi.org/10.1038/s41598-024-56415-5
3. E. Erak, L.D. Oliveira, A.A. Mendes, O. Dairo, O. Ertunc, I. Kulac, J.B.-D. Valle, T. Jones, J.L. Hicks, S. Glavaris, G. Guner, I.D. Vidal, M. Markowski, C. De La Calle, B.J. Trock, A. Meena, U. Joshi, C. Kondragunta, S. Bonthu, N. Singhal, A.M. De Marzo, T.L. Lotan, Predicting prostate cancer molecular subtype with deep learning on histopathologic images. Mod. Pathol. **36**, 100247 (2023). https://doi.org/10.1016/j.modpat.2023.100247
4. N. Rabilloud, P. Allaume, O. Acosta, R. De Crevoisier, R. Bourgade, D. Loussouarn, N. Rioux-Leclercq, Z.E. Khene, R. Mathieu, K. Bensalah, T. Pecot, S.F. Kammerer-Jacquet, Deep learning methodologies applied to digital pathology in prostate cancer: a systematic review. Diagnostics **13**, 2676 (2023). https://doi.org/10.3390/diagnostics13162676
5. W. Bulten, K. Kartasalo, P.-H.C. Chen, P. Ström, H. Pinckaers, K. Nagpal, Y. Cai, D.F. Steiner, H. Van Boven, R. Vink, C.H.-V. De Kaa, J. Van Der Laak, M.B. Amin, A.J. Evans, T. Van Der Kwast, R. Allan, P.A. Humphrey, H. Grönberg, H. Samaratunga, B. Delahunt, T. Tsuzuki, T. Häkkinen, L. Egevad, M. Demkin, S. Dane, F. Tan, M. Valkonen, G.S. Corrado, L. Peng, C.H. Mermel, P. Ruusuvuori, G. Litjens, M. Eklund, A. Brilhante, A. Çakır, X. Farré, K. Geronatsiou, V. Molinié, G. Pereira, P. Roy, G. Saile, P.G.O. Salles, E. Schaafsma, J. Tschui, J. Billoch-Lima, E.M. Pereira, M. Zhou,

S. He, S. Song, Q. Sun, H. Yoshihara, T. Yamaguchi, K. Ono, T. Shen, J. Ji, A. Roussel, K. Zhou, T. Chai, N. Weng, D. Grechka, M.V. Shugaev, R. Kiminya, V. Kovalev, D. Voynov, V. Malyshev, E. Lapo, M. Campos, N. Ota, S. Yamaoka, Y. Fujimoto, K. Yoshioka, J. Juvonen, M. Tukiainen, A. Karlsson, R. Guo, C.L. Hsieh, I. Zubarev, H.S.T. Bukhar, W. Li, J. Li, W. Speier, C. Arnold, K. Kim, B. Bae, Y.W. Kim, H.S. Lee, J. Park, Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat. Med. **28**, 154–163 (2022). https://doi.org/10.1038/s41591-021-01620-2

6. L. Zhu, J. Pan, W. Mou, L. Deng, Y. Zhu, Y. Wang, G. Pareek, E. Hyams, B.A. Carneiro, M.J. Hadfield, W.S. El-Deiry, T. Yang, T. Tan, T. Tong, N. Ta, Y. Zhu, Y. Gao, Y. Lai, L. Cheng, R. Chen, W. Xue, Harnessing artificial intelligence for prostate cancer management. Cell Rep. Med. **5**, 101506 (2024). https://doi.org/10.1016/j.xcrm.2024.101506
7. D. Petríková, I. Cimrák, Survey of recent deep neural networks with strong annotated supervision in histopathology. Computation **11**, 81 (2023). https://doi.org/10.3390/computation11040081
8. W. Mi, J. Li, Y. Guo, X. Ren, Z. Liang, T. Zhang, H. Zou, Deep learning-based multi-class classification of breast digital pathology images. Cancer Manag. Res. **13**, 4605–4617 (2021). https://doi.org/10.2147/cmar.s312608
9. J. Huang, L. Mei, M. Long, Y. Liu, W. Sun, X. Li, H. Shen, F. Zhou, X. Ruan, D. Wang, S. Wang, T. Hu, C. Lei, BM-Net: CNN-based MobileNet-V3 and bilinear structure for breast cancer detection in whole slide images. Bioengineering **9**, 261 (2022). https://doi.org/10.3390/bioengineering9060261
10. S. Abousamra, R. Gupta, L. Hou, R. Batiste, T. Zhao, A. Shankar, A. Rao, C. Chen, D. Samaras, T. Kurc, J. Saltz, Deep learning-based mapping of tumor infiltrating lymphocytes in whole slide images of 23 types of cancer. Front. Oncol. (2022). https://doi.org/10.3389/fonc.2021.806603
11. M. Abdolahi, M. Salehi, I. Shokatian, R. Reiazi, Artificial intelligence in automatic classification of invasive ductal carcinoma breast cancer in digital pathology images. Med. J. Islam Repub. Iran (2020). https://doi.org/10.47176/mjiri.34.140
12. F. Wessels, M. Schmitt, E. Krieghoff-Henning, J.N. Kather, M. Nientiedt, M.C. Kriegmair, T.S. Worst, M. Neuberger, M. Steeg, Z.V. Popovic, T. Gaiser, C. Von Kalle, J.S. Utikal, S. Fröhling, M.S. Michel, P. Nuhn, T.J. Brinker, Deep learning can predict survival directly from histology in clear cell renal cell carcinoma. PLoS ONE **17**, e0272656 (2022). https://doi.org/10.1371/journal.pone.0272656
13. L. Jin, F. Shi, Q. Chun, H. Chen, Y. Ma, S. Wu, N.U.F. Hameed, C. Mei, J. Lu, J. Zhang, A. Aibaidula, D. Shen, J. Wu, Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. Neuro Oncol. **23**, 44–52 (2020). https://doi.org/10.1093/neuonc/noaa163
14. O. Attallah, MB-AI-His: histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. Diagnostics **11**, 359 (2021). https://doi.org/10.3390/diagnostics11020359
15. O. Attallah, CoMB-Deep: composite deep learning-based pipeline for classifying childhood medulloblastoma and its classes. Front. Neuroinform. (2021). https://doi.org/10.3389/fninf.2021.663592
16. H. Fu, W. Mi, B. Pan, Y. Guo, J. Li, R. Xu, J. Zheng, C. Zou, T. Zhang, Z. Liang, J. Zou, H. Zou, Automatic pancreatic ductal adenocarcinoma detection in whole slide images using deep convolutional neural networks. Front. Oncol. (2021). https://doi.org/10.3389/fonc.2021.665929
17. H. Yang, L. Chen, Z. Cheng, M. Yang, J. Wang, C. Lin, Y. Wang, L. Huang, Y. Chen, S. Peng, Z. Ke, W. Li, Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. BMC Med. (2021). https://doi.org/10.1186/s12916-021-01953-2
18. A. Raczkowska, I. Paśnik, M. Kukiełka, M. Nicoś, M.A. Budzinska, T. Kucharczyk, J. Szumiło, P. Krawczyk, N. Crosetto, E. Szczurek, Deep learning-based tumor microenvironment segmentation is predictive of tumor mutations and patient survival in non-small-cell lung cancer. BMC Cancer (2022). https://doi.org/10.1186/s12885-022-10081-w
19. Z. Hameed, B. Garcia-Zapirain, J.J. Aguirre, M.A. Isaza-Ruget, Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. Sci. Rep. (2022). https://doi.org/10.1038/s41598-022-19278-2
20. N. Cheng, Y. Ren, J. Zhou, Y. Zhang, D. Wang, X. Zhang, B. Chen, F. Liu, J. Lv, Q. Cao, S. Chen, H. Du, D. Hui, Z. Weng, Q. Liang, B. Su, L. Tang, L. Han, J. Chen, C. Shao, Deep learning-based classification of hepatocellular nodular lesions on whole-slide histopathologic images. Gastroenterology **162**, 1948-1961.e7 (2022). https://doi.org/10.1053/j.gastro.2022.02.025
21. X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, X. Han, Transformer-based unsupervised contrastive learning for histopathological image classification. Med. Image Anal. **81**, 102559 (2022). https://doi.org/10.1016/j.media.2022.102559
22. N. Elazab, W.A. Gab-Allah, M. Elmogy, A multi-class brain tumor grading system based on histopathological images using a hybrid YOLO and RESNET networks. Sci. Rep. (2024). https://doi.org/10.1038/s41598-024-54864-6
23. R. Yan, Z. Yang, J. Li, C. Zheng, F. Zhang, Divide-and-attention network for HE-stained pathological image classification. Biology **11**, 982 (2022). https://doi.org/10.3390/biology11070982
24. R. Yan, F. Ren, J. Li, X. Rao, Z. Lv, C. Zheng, F. Zhang, Nuclei-guided network for breast cancer grading in HE-stained pathological images. Sensors **22**, 4061 (2022). https://doi.org/10.3390/s22114061
25. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016)
26. S.J. Pan, Q. Yang, A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**, 1345–1359 (2009). https://doi.org/10.1109/tkde.2009.191

27. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269, 2017. https://doi.org/10.1109/cvpr.2017.243
28. K. He K, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016. https://doi.org/10.1109/CVPR.2016.90