RESEARCH ARTICLE

Deep Learning Approaches for Morphological Classification of Intestinal Organoids

GIOVANNI CICCERI [ID] [1,*], **SEBASTIANO DI BELLA** [ID] [2,*], **SIMONE DI FRANCO** [ID] [2], **GIORGIO STASSI** [2], **MATILDE TODARO** [3], **AND SALVATORE VITABILE** [ID] [1]

[1] Department of Biomedicine, Neuroscience, and Advanced Diagnostics (BiND), University of Palermo, 90127 Palermo, Italy
[2] Department of Precision Medicine in Medical, Surgical, and Critical Care (MePreCC), University of Palermo, 90127 Palermo, Italy
[3] Department of Health Promotion Sciences, Internal Medicine, and Medical Specialties (ProMISE), University of Palermo, 90127 Palermo, Italy

Corresponding author: Giovanni Cicceri (giovanni.cicceri@unipa.it)

ABSTRACT Organoids, derived from primary donor or stem cells, closely replicate the composition and function of their *in vivo* counterparts. This quality makes them a reliable model for validating hypotheses on disease-related biological processes and mechanisms. To date, the classification of organoids is performed manually by microscope and, therefore, in a data-driven application, is time-consuming, inaccurate, and difficult to morphological analysis process. The use of deep learning (DL) in organoid image analysis becomes crucial to handle complexity, variability, and large amounts of data efficiently and accurately, overcoming the limitations of traditional image processing approaches. In this paper, five CNN-based DL models such as MobileNet, DenseNet, ResNet, Inception, VGG, and the very recent Vision Transformers (ViT) were analyzed using a publicly available dataset for the morphological classification of intestinal organoids. Additionally, traditional ML models, such as SVM and RF, were tested for comparison using a feature set similar to conventional image processing tools. The systematic performance evaluation is designed to guide users in choosing the most suitable model for processing organoid images. Among all models, ViT achieved the highest accuracy of 86.95%, demonstrating its effectiveness in organoid classification. Inception and DenseNet also exhibited strong performance, with accuracy values of 86.10% and 86.47%, respectively. Rather, SVM and RF performed significantly worse, showing an accuracy approximately 20% lower than the selected DL models. Considering efficiency, ViT had the highest accuracy but required more resources (0.0437 sec/image, 343 MB), while MobileNet, the lightest model (35.6 MB), had the fastest inference time (0.0063 sec/image). The findings highlight the potential of DL models in enhancing the accuracy of organoid classification while emphasizing the importance of balancing performance with computational efficiency for real-time applications.

INDEX TERMS Deep learning, image classification, intestinal organoids, stem cells, cancer.

I. INTRODUCTION

Cancer is one of the leading causes of death worldwide, especially the incidence of colorectal cancer (CRC) is the highest in developed countries; based on ageing, population growth, and human development, the global number of new

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

CRC cases is estimated to be about 3 million in the next 15 years [1], [2]. Organoid culture provides a physiologically relevant model that maintains the cellular composition and genetic profiles of the original tissue over multiple passages [3], [4]. Since organoids mimic a solid tumor *in vivo*, they are a useful approach as a tool to study tumor biology, metabolism, resistance and metastasis [5], [6]. Organoids can be generated from almost any tissue from different cancer

patients, from the primary tumor or metastasis to ascites and pleural fluid. Unlike whole tissue samples, they do not require intensive tissue digestion and result in a miniature organ that is reconstituted and incorporated into the extracellular matrix in an attempt to maintain the constitution of the tumor microenvironment. Therefore, organoids reproduce the architecture of the original tumor, and even long-term culture maintains gene stability and preserves the mutational profile of the original tumors [7], [8]. Organoids have played an important role in current cancer research, especially in the study of the tumor development process, improving the efficiency of tumor therapy and tumor heterogeneity [9]. Classification and quantitative analysis of organoids are instrumental in studying the physical and molecular characteristics of a complex phenomenon such as CRC.

In recent years, organoids have been widely used in cancer research [10], [11], [12], [13], [14], in fact, they mimic and recapitulate several functional and structural aspects of a number of carcinomas, for example, intestinal organoids can develop crypt-like structures similar to those in the native colon, these structures are fundamental for the regeneration of the epithelium *in vivo*. Organoids are self-organizing and self-renewing structures in 3D culture *in vitro* that are embedded in a biological matrix to create an extracellular environment that provides structural support and essential growth factors [15], [16].

Although studying the physical and molecular characteristics of complex phenomena such as CRC through the classification and quantitative analysis of organoids is beneficial, the study of metrics such as changes in morphology, number, and size can be challenging and time-consuming in large data-driven experiments [17], [18]. This is due to the need for manual detection, labeling, and analysis of numerous images. Consequently, developing an automated pipeline for processing organoid images is difficult, as it involves addressing issues such as overlap, collapse, and out-of-focus problems. These issues, usually encountered in live imaging modality, can lead to overlapping structures that are difficult to distinguish individually and to the structural collapse of organoids, potentially distorting their morphology by making accurate analysis challenging. Organoid morphology offers valuable insights into colon physiology and pathophysiology *in vivo*. For instance, opacity and budding are key indicators used to assess the maturity of colon organoids, while cystic organoids, which possess thinner walls, are more transparent, whereas colon organoids tend to be more opaque.

The challenges inherent in organoid image analysis, such as the detection of complex morphologies, variability between samples, and focus issues, make traditional approaches to image processing suboptimal. In such a context, the move to Machine Learning (ML) and the subset Deep Learning (DL) is not just a trend but a necessity, as these data-driven approaches can leverage advanced models to overcome the traditional limitations of image processing, enabling more nuanced and automated analysis of morphological complexities.

This work aims to provide the research community with valuable insights into the performance and applicability of various DL models for the automatic classification of intestinal organoids. Specifically, the experimental design incorporated six DL techniques for classification tasks, including the most used Convolutional Neural Networks (CNN)-based approaches like MobileNetV2, DenseNet169, ResNet50v2, InceptionV3, and VGG16, along with the cutting-edge DL methods, such as Vision Transformers (ViT), which have recently gained significant traction in the field of image classification. All experiments were conducted using a publicly available dataset (described in Section III). To provide a more comprehensive evaluation, comparisons with conventional ML models, such as Support Vector Machines (SVM) and Random Forest (RF), are included, utilizing a feature set similar to those employed in conventional image processing tools. The results indicate that ViT model, albeit slightly, outperformed traditional CNN-based models in terms of accuracy, precision, recall, specificity, and F1-score. However, these DL-based computational models offer a more efficient alternative to labor-intensive and time-consuming high-throughput experimental techniques currently used to study organoids.

This research marks a significant shift in focus from the development of algorithms and models for organoid analysis to the comprehensive application and evaluation of existing DL models for organoid classification, thereby not only establishing a benchmark for DL algorithms in this domain but also significantly contributing to the advancement of automated analysis methodologies and technologies in the field of cell culture models. In more specific terms, the main contributions of this work are *five-fold*:

i) This work expands upon existing DL models by developing or adapting algorithms capable of identifying intestinal organoids from low-resolution transmitted-light images across a broader range of tissues;

ii) A high-level workflow is proposed for evaluating DL models based on three critical aspects: *model configuration*, *model accuracy*, and *model efficiency*. This workflow is designed to serve as a benchmark for future DL model assessments in organoid classification;

iii) The dataset obtained from the image extraction algorithm has been released to the research community for further investigation;

iv) A systematic comparison of various DL models is provided to help researchers meticulously assess the strengths and weaknesses of each DL model. Additionally, a comparative analysis of traditional ML models is included to highlight the advantages of DL approaches over conventional methods. This comparison serves as a decision-making tool to guide researchers toward the most suitable model for accurate intestinal organoid classification;

v) The analysis identifies the ViT as the model with the best overall performance in organoid classification. This

finding underscores the potential of ViT as a leading model in the field and sets a precedent for its application to similar tasks.

These contributions collectively enhance the understanding and development of automated analysis methodologies in the field of intestinal organoid classification, paving the way for future exploration of large-scale DL applications in cell culture models. The rest of the paper is structured as follows. Section II provides an overview of the existing approaches in the field of computational biology. Section III details the materials and the methodology employed in this work, outlining the complete handling of the data, followed by an in-depth discussion of the proposed models. Section IV presents the experimental results by rating the performance of all the proposed models. Section V examines the implications of the results, comparing the performance of the proposed models and highlighting their advantages and limitations. Finally, Section VI concludes the work with a summary of the key findings and future research directions.

II. RELATED WORKS

The evolving field of computational biology and biomedical informatics has brought forward advanced algorithms capable of dissecting the intricate data derived from organoid studies [19], [20]. To such evidence, authors in [21] simplify the complexity of genetic and cellular organization by analyzing over 200.000 live cells using the WTC-11 hiPSC Single-Cell Image Dataset v1, focusing on key cellular structures to map distinct organelles and functional machines. This approach opened a new pathway for biological data exploration through novel data-driven approaches, enabling human-interpretable, dimensionally reduced quantitative measurements while embracing cell-to-cell variability observed in normal populations. In [22], the authors introduce a data-driven approach to address the challenge of intratumor phenotypic and functional heterogeneity in acute myeloid leukemia (AML) by using mass cytometry to profile surface and intracellular signalling proteins across millions of cells. Using PhenoGraph, a partitional algorithmic tool developed to define phenotypes in high-dimensional single-cell data, the research uncovers that surface phenotypes of leukemic blasts do not reliably indicate their intracellular state. This methodology enables the identification of a signaling-based cellular phenotype, leading to the discovery of a gene expression signature predictive of survival in AML, offering new insights into its pathophysiology and enhancing large-scale analysis of single-cell heterogeneity.

Recent breakthroughs in automated microscopy have transformed high-throughput cell biology, enabling complex image-based screenings that are better navigated through ML methods [23]. These approaches leverage data structures and biologist annotations to develop versatile analytical models [24], [25]. Integrating high-throughput imaging and ML approaches has revolutionized phenotypic profiling in

drug discovery, offering new insights into the mechanisms, efficacy, and toxicity of small molecules [26].

Numerous studies have applied ML approaches to organoid research. Authors in [27] utilized linear support vector machines to distinguish correctly segmented organoids from those with artefacts, using features like PhenoGraph analysis and Zernike polynomials. Although ML techniques have proven useful, DL models, particularly convolutional neural networks (CNNs), often outperform traditional ML in image analysis and pattern recognition tasks. Neural networks excel at analyzing complex morphologies, providing a scalable solution suitable for organoid analysis. For instance, OrgaQuant [28] and OrganoID [29] were developed to automate the detection and tracking of organoids, facilitating high-throughput studies in drug discovery and personalized medicine.

In addition, several organoid image analysis software and tools have been proposed over the years, ranging from visualizers like Fiji/ImageJ [30], Napari [31] or QuPath [32] to comprehensive software that use specific algorithms/models for advanced tasks such as automated organoid segmentation and feature extraction, like MOrgAna [33], OrganoidTracker [34], ilastik [35] and CellProfiler [36]. These tools, however, often require extensive domain expertise to fine-tune analysis pipelines for complex morphological patterns.

Several more recent studies have introduced DL approaches for organoid analysis. Authors in [37] harness electron microscopy (EM) to provide an intricate view of cellular morphology, surpassing traditional analysis limitations with a novel unsupervised neural network approach for learning morphological features directly from 3D EM data. In [38], the authors leverage the unsupervised geometric DL approach to capture 3D cell shape representations from microscopy images of metastatic melanoma cells, addressing the need for comprehensive 3D shape quantification in medical research. In [39], the authors employed high-throughput, image-based profiling (by using a deep convolutional neural network - DCNN) to analyze over 5 million individual CRC organoids treated with more than 500 small molecules, revealing key axes of morphological variation influenced by IGF1 receptor signalling and LGR5+ stemness.

Authors in [40] introduced an organoid image dataset and an innovative deep neural network for automated detection and tracking of organoids, setting a new standard for efficiency in organoid research methodologies. In [41], a DL model was developed using bright-field images to predict retinal differentiation in stem cell-derived organoids, outperforming experts with an 84% accuracy. Authors in [42] explored how CNNs can offer faster and more accurate approaches to organoid morphology. In [43], the authors introduced a morphological screening method using deep neural networks to identify drugs reversing epithelial-mesenchymal transition in mammary

tumor organoids. Recent advances in computer vision have led to more sophisticated and accurate image classification and detection procedures on organoids [44], [45]. The recent emergence of Vision Transformer (ViT) models marks a significant change from traditional CNNs, however, which have long dominated this field. Authors in [46] developed *TransOrga* method, exploiting the transformer architecture, offering an advanced tool for noninvasive segmentation of organoids. In [47], the authors developed the RDAU-Net model, an advanced DL tool designed to enhance the accuracy and efficiency of organoid-based drug screenings using improved dynamic convolution and attention mechanisms. Despite these advancements, a comprehensive benchmarking of DL models for organoid classification remains critical. Benchmarking serves multiple purposes: it helps identify the most effective models, ensures consistent and accurate analysis, and highlights the strengths and limitations of different approaches. This process is vital for several reasons. Firstly, it standardizes the evaluation of DL models, facilitating their adoption in diverse research settings. Secondly, it can enhances the reliability of phenotypic profiling, which is crucial for understanding disease mechanisms and responses to treatments. Lastly, it can accelerate the drug discovery process by providing robust tools for high-throughput screening of organoids.

III. MATERIALS AND METHODS

Figure 1 presents the high-level workflow aimed at image classification of organoids. It consists of five main stages: i) *Data source*, ii) *Data pre-processing and labeling*, *iii) Datasets partition*, iv) *Model selection*, and v) *Model evaluation*. The details and complexities of each step in such workflow are explained in the following sections.

A. DATA SOURCE

As a data source, a well-established dataset specifically designed and validated for the organoid detection task by Tellu [48] was utilized. This dataset, developed by Domènech-Moreno et al., is publicly available on Zenodo repository (YOLO format) at https://doi.org/10.5281/zenodo.6768583. It was generated using an EVOS FL microscope (Thermo Fisher) with a 4x objective and consists of 840 transmitted-light microscopy images of intestinal organoids taken during different stages. The authors used the Python library *lablImg* to annotate images in *four* classes, in particular, they coded with *Class0* for *organoid0* the class "Cyst" that represents a very early stage, small, cystic, non-budding organoids with thick wall, with *Class1* for *organoid1* the class "Ealy organoid" indicating early, budding organoids with 1-2 crypts (buds), with *Class2* for *organoid3* the class "Late organoid" containing large, budding organoids with 3 or more crypt units and finally *Class3* for *spheroid* for a class "Spheroid" showing large, circular, thin-welled organoids representing e.g. fetal, regenerating, tumorigenic or hyperstimulated organoids. The entire original dataset (described in Table 1) contains

23.065 annotations, however, the four classes are not equally distributed because all organoids start from a cystic shape. In detail, the dataset was composed in *Class0* for 52% with 11.922 instances, *Class1* for 24% with 5.510 instances, *Class2* for 14% with 3.367 instances, and then *Class3* with 10% with 2.266 instances.

B. DATA PRE-PROCESSING AND LABELING

To evaluate the performance of the models, a comprehensive data pre-processing and labeling strategy was implemented, utilizing a custom Python script available in the GitHub repository (extract.py file in "Python scripts" folder) at the following URL: https://github.com/gcicceri/Organoids-Classification-). This script was specifically designed to handle images of intestinal organoids by cropping each image to a standard size of 224×224 pixels, which is a common input size for DL models. [49]. This standardization to a fixed resolution was essential to maintaining consistency across models, reducing variability in feature extraction, and ensuring compatibility with the pre-trained DL architectures.

In detail, the script operates by reading the bounding box coordinates from the original annotation files to extract the relevant image sections and associate them with their respective labels. During this process, two annotations within Class0 were identified as inconsistent, as the specified coordinate pairs did not produce a valid 224×224 clipping mask.

Upon closer inspection, it was evident that these bounding boxes corresponded to regions that did not contain any organoids. Specifically, Image 18 and Image 710 were identified as problematic. These two images can be viewed in the GitHub repository, where the problematic regions have been highlighted as areas that do not enclose any organoids.

As a result, these images were excluded from subsequent partitioning and, consequently, from the final dataset. Specifically, this exclusion ensured that only correctly labeled and properly segmented organoid images were used for training, thus preventing potential sources of noise that could mislead the learning process. Nevertheless, their removal had a negligible impact on overall classification performance. The resulting dataset comprises 23,063 images, each associated with a class label (Class0, Class1, Class2, Class3), and was used for training and testing the models.

Figure 2 shows a representative subset of the large set of images of the four classes, specifically selected to demonstrate the different morphology of organoids in the distinguished classes.

C. DATASETS PARTITION

To facilitate more efficacious learning and testing processes, the dataset (shown in Table 2) was partitioned into three subsets: i) *training set* composed of 18.537 images, ii) *validation set* containing 2.058 images and iii) *test set* including 2.468 images. This structured partitioning is a pivotal element in the performance assessment of DL
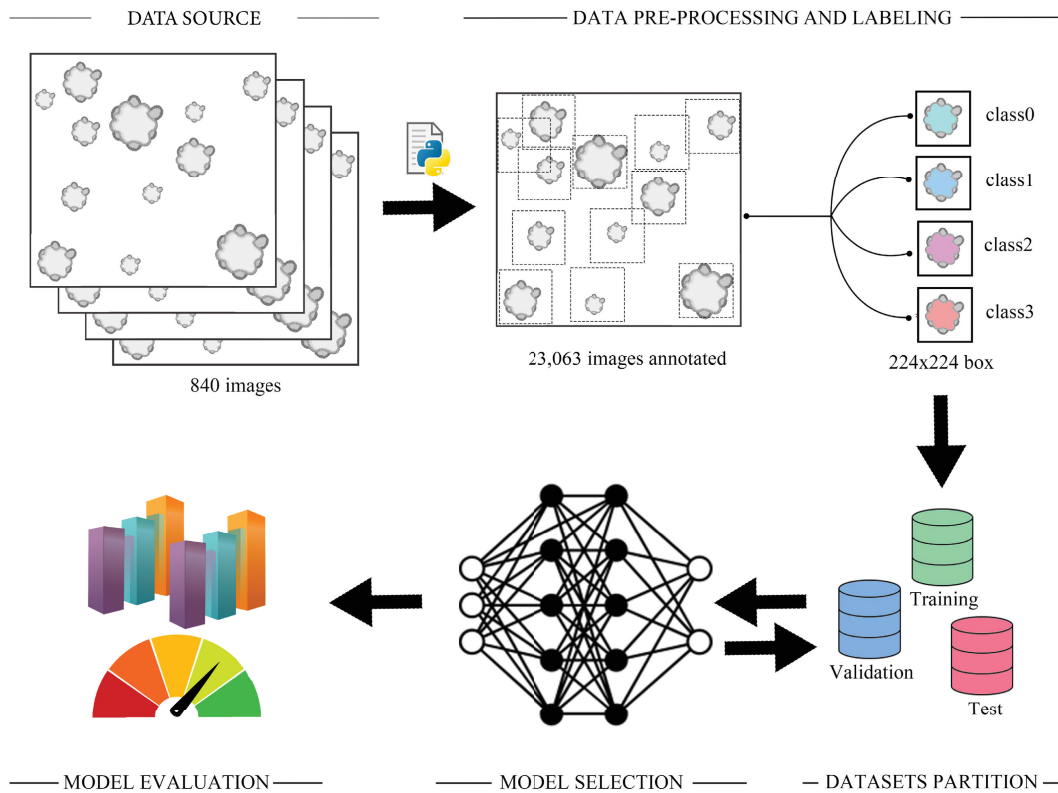
FIGURE 1. The proposed High-Level workflow.

TABLE 1. Data source distribution for *organoid0* (Class0), *organoid1* (Class1), *organoid3* (Class2) and *spheroid* (Class3).

	Class0	Class1	Class2	Class3	Total
Data source [48]	11.922 (52%)	5.510 (24%)	3.367 (14%)	2.266 (10%)	**23.065**

TABLE 2. Images annotated distribution in a train, validation, and test set for *organoid0* (Class0), *organoid1* (Class1), *organoid3* (Class2) and *spheroid* (Class3).

Data	Class0	Class1	Class2	Class3	Total
Train	9.564 (51.60%)	4.466 (24.10%)	2.670 (14.40%)	1.837 (9.90%)	18.537
Validation	1.062 (51.60%)	496 (24.10%)	296 (14.39%)	204 (9.91%)	2.058
Test	1.294 (52.43%)	548 (22.20%)	401 (16.25%)	225 (9.12%)	2.468

architectures. The training dataset is utilized for the initial fitting of the model, enabling it to learn from the data. Subsequently, the validation dataset plays an integral role in hyperparameters optimization, facilitating the refinement of the model. Finally, the test dataset serves as the benchmark for evaluating the efficacy of the models, providing an unbiased assessment of their predictive capabilities. The final dataset has been made publicly available to the research community. It can be accessed through the Zenodo repository (see https://doi.org/10.5281/zenodo.14725323).

D. MODEL SELECTION

To accurately classify organoids, a set of six state-of-the-art DL models was employed. Specifically, five recent Convolutional Neural Network (CNN)-based DL models like MobileNet, DenseNet, ResNet, Inception, VGG, and the very

recent Vision Transformers (ViT). These selected DL models represent distinct architectural paradigms, enabling a comprehensive evaluation of organoid classification strategies. Each model excels in feature extraction, computational efficiency, or classification accuracy, offering a complementary assessment. A concise overview of each model's role in the multi-class classification task follows. To enhance comparison and validate DL efficacy, conventional ML models (SVM and RF) were also included using manually extracted features. The following presents a detailed summary of each selected DL/ML model.

1) DenseNet169

DenseNet169 is a variant of the Dense Convolutional Network (DenseNet), a type of CNN introduced by Gao Huang et al. [50], [51], and known for its dense
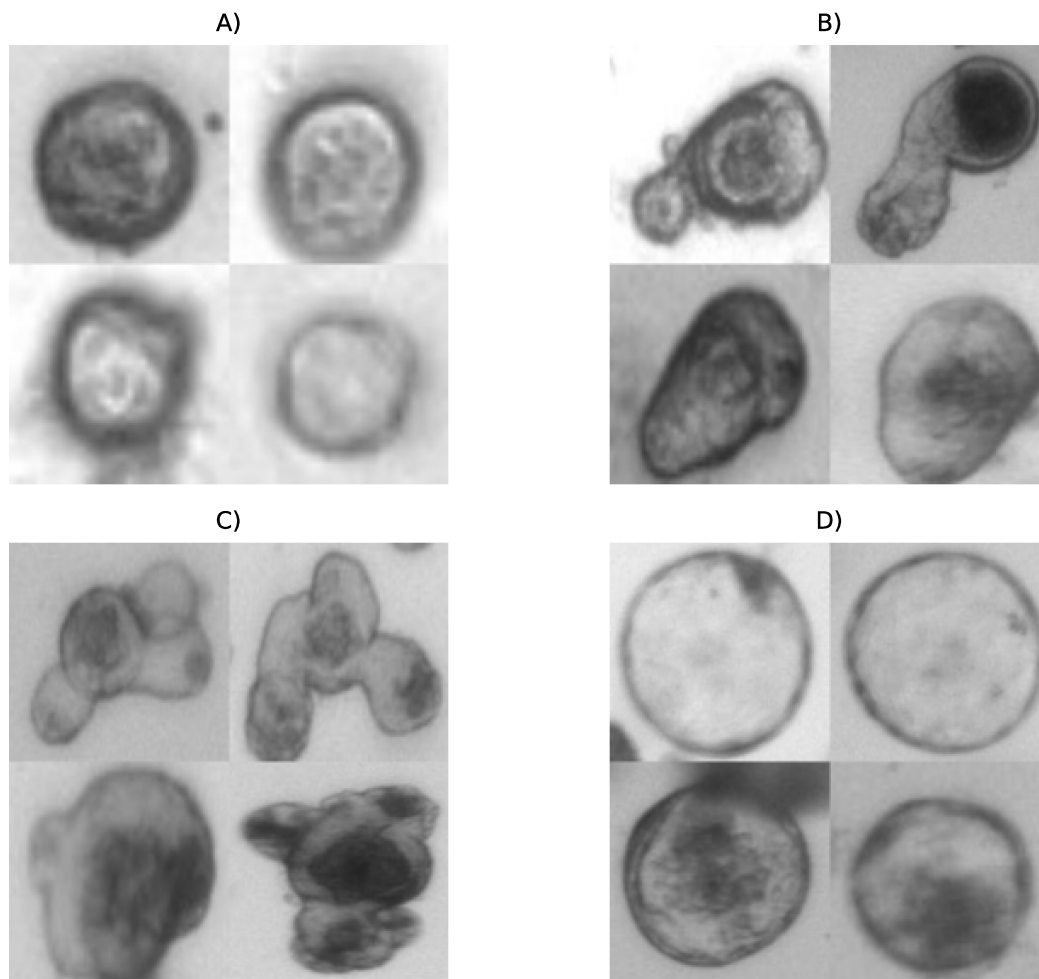
FIGURE 2. Sample images of organoid morphology across the four different classes: A) Class0 - *organoid0*. B) Class1 - *organoid1*. C) Class2 - *organoid3*. D) Class3 - *spheroid*.

connectivity pattern. DenseNet169, commonly used for image classification tasks, refers to a DenseNet architecture with 169 layers if one counts each of normalization, nonlinearity, pooling, and convolution as separate layers, and has the advantage of requiring fewer parameters than other CNN architectures, often achieving better performance. Unlike traditional CNNs, where each layer is only connected to the next layer, in DenseNet, for each layer, the feature maps of all preceding layers are used as inputs, and its own feature maps are used as inputs into all subsequent layers. This helps with gradient flow and allows for more efficient use of parameters. This design improves information flow and feature propagation, enabling better performance even with fewer layers. Its densely connected layers improve feature reuse and gradient flow, enhancing the model's ability to capture fine-grained morphological details in organoid structures.

2) ResNet50v2

ResNet50v2 is a variant of the ResNet (Residual Network) architecture, the family originally introduced by Kaiming

He and his team in 2015 [52] for practical applications such as medical image analysis, autonomous driving, and object detection. It is a CNN-based model designed to tackle the challenges associated with training very deep networks. ResNet50 uses "residual connections", which bypass one or more layers and are added back to the output. This idea enables the training of much deeper networks (even hundreds or thousands of layers) without severe degradation of training accuracy. The "v2" in ResNet50v2 [53], proposed in this study as the comparison model, refers to the second version of the original ResNet50. The difference between original ResNet v1 and ResNet v2 is in the block design. Specifically, in v1, the skip connection skips over the weight layers. In v2, the order of operations in the residual block is changed to start with batch normalization and rectified linear unit (ReLU) before convolution, which is claimed to give better accuracy. The architecture of ResNet50 consists of 50 layers stacked neatly in an efficient architecture, including convolutional layers, batch normalization, ReLU activation functions, and fully connected layers. Each "residual block" of the architecture consists of a set of layers whose output is

added to the input, making up the "residual connection". By utilizing residual learning, ResNet50v2 mitigates vanishing gradient issues and preserves both low- and high-level spatial features, improving classification accuracy, particularly in ambiguous cases, such as organoid morphologies with overlapping or unclear structural characteristics. ResNet50 revolutionized the approach to DL models, paving the way for later architectures such as DenseNet (outlined above), and solving major challenges in training deep networks.

3) MobileNetV2

MobileNetV2 is another neural network model used in this study. MobileNetV2, introduced in [54], designed specifically for mobile and other low-computational-power devices, where computational resources are limited, represents an evolution of the original MobileNet model [55] and offers improved performance and efficiency. Its optimized characteristics make it particularly suitable for real-time organoid classification and deployment in resource-constrained environments. MobileNetV2 architecture is based on the "inverted residuals" procedure, unlike traditional residual blocks, in which the channel size is first reduced and then expanded, MobileNetV2 does the opposite, that is, it first expands the size (using a 1×1 convolution), then applies the convolution in-depth, and finally reduces the size. Moreover, after the depthwise convolution, instead of applying a ReLU activation before the pointwise convolution, the model maintains a linear connection, helping to preserve the information. MobileNetV2 also employs depthwise separable convolutions to reduce the number of parameters and computations. This involves separating the convolution operation into a depthwise convolution followed by a pointwise convolution.

4) InceptionV3

The Inception model, introduced by Google researchers, has emerged as one of the most innovative models computer vision field, radically challenging traditional standards CNNs [56]. InceptionV3 is the third iteration of the Inception architecture [57] and includes improvements and refinements over the previous versions. The architecture of InceptionV3 introduced several advancements, such as factorized convolutions and label smoothing, which led to better performance and efficiency, particularly in the context of image recognition tasks. Leveraging multi-scale feature extraction, this model effectively handles the heterogeneity of organoid shapes, making it a robust choice for complex image classification tasks, while its regularization techniques contribute to improved generalization and training stability. Specifically, key aspects of InceptionV3 involve factorization into smaller convolutions, where large convolutions are factorized into smaller, more manageable operations, and spatial factorization into asymmetric convolutions, in which InceptionV3 often uses asymmetric convolutions, such as 1×3 followed by 3×1 convolutions, instead of larger square

convolutions. In addition, the net employs auxiliary classifiers (with batch normalization) during training, improving faster convergence, providing additional regularization, and helping to solve the vanishing gradient problem, making it efficient and scalable in terms of both computational resources and performance.

5) VGG16

VGG16 is a variant of the VGG model, developed by the Visual Geometry Group (VGG) at the University of Oxford in 2014 [58]. Its DL architecture quickly garnered attention for its excellent performance in the ImageNet Large-Scale Visual Recognition Challenge. As one of the earliest deep CNN architectures, VGG16 serves as a crucial benchmark in organoid classification, offering a straightforward and interpretable structure for feature extraction. VGG16's architecture consists of 16 weight layers and consists of a series of convolutional layers with small 3×3 receptive fields, followed by max-pooling layers for spatial down-sampling. Several layers are connected in this fashion, culminating in a softmax output layer. This uniformity provides simplicity and depth, which were significant factors in the network's successful performance. Its structured feature hierarchy enables effective extraction of morphological patterns in organoid images, making it a valuable baseline for comparison with more complex architectures. Despite VGG16 has a very high number of parameters (138 million), which makes it computationally intensive and memory demanding its uniform structure has made it a popular choice for transfer learning in various computer vision tasks, and its impact continues to influence subsequent models, making it a foundational architecture in DL.

6) VISION TRANSFORMERS

The Vision Transformers (ViT) is an innovative approach to computer vision applications that adapts the transformer architecture by simulating neural processing in the human brain [59]. It had previously been highly successful in natural language processing tasks, showing the full power of transformers for image classification tasks, a domain traditionally dominated by CNNs, and for this reason, compared in this work with the latter. Unlike traditional CNNs, which process images using a hierarchical approach, ViT begins by splitting the input image into fixed-size patches (in pixels). The tokenized and embedded patches are then fed into a series of transformer encoder layers. The transformer processes these tokens in parallel, attending to different patches based on their content and relative positions, similar to what is done in natural language processing (NLP) transformers. Specifically, in the learning phase, ViT dissects each image into smaller patches, converts them into vectors, and processes them through layers employing the Multi-Head Attention mechanism [60]. This approach adeptly captures local features while maintaining an awareness of the overall

image context, enabling a more holistic understanding of organoid morphology. ViT's efficacy has been validated in multiple clinical use cases, particularly for image classification tasks [61], [62], [63]. Due to its demonstrated effectiveness, ViT was chosen as the reference model for this multi-class organoid classification task.

7) SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) model was used as basic comparison with the proposed DL-based models. SVM is a supervised ML model designed for both binary and multi-classification tasks [64]. It works well for both linearly and non-linearly separable data, and it has been widely applied to various kinds of problems, including image classification. SVM model aims to find a hyperplane that optimally separates the different classes while maximizing the margin between them. Unlike CNNs, SVM do not automatically handle the spatial hierarchies of images, so to use SVM for traditional image classification problems, it is necessary to extract image feature descriptors such as color histograms, textures, shape, and so on. One of the most important aspects to emphasize about SVM is the ability to use kernel functions to transform the feature space into higher dimensions in which classes are linearly separable.

8) RANDOM FOREST

Random Forest (RF) is an additional conventional ML model. It was used as a further comparison to the DL-based models for the organoid classification task. RF [65] is an ensemble learning algorithm that constructs a multitude of decision trees during the training process. Each tree is built using a sample drawn from the original dataset, which means that some data points may be repeated while others are omitted in each tree's training. Additionally, at each decision node, the algorithm selects a random subset of features to determine the best split. This feature selection process plays a crucial role in mitigating overfitting, as it prevents any single tree from being overly dependent on any particular set of features. By limiting the number of features considered at each split, the model enhances its generalization capabilities, thereby reducing the risk of overfitting and improving overall performance.

E. IMAGE FEATURE EXTRACTION FOR ML MODELS

Feature extraction is a pivotal step in image processing, converting raw data into meaningful quantitative information. In this study, feature extraction played a key role in distinguishing organoid classes based on their characteristics and was instrumental in training the two proposed ML models. The selected features were carefully chosen to align with those used in traditional tools such as ilastik, ensuring consistency and relevance in the comparative analysis. The extracted features (detailed in Table 3) were grouped into three primary categories: *texture*, *shape*, and *color*. Texture features captured the complex patterns within organoid

structures. Shape features quantified the geometry and size of the organoids. Color features provided a quantitative view of the image's color properties. Collectively, the extracted features (in total 28) offered a robust approach for the detailed characterization of organoid images, ensuring that the organoids were classified not only by developmental stage but also by their morphological traits. Segmentation was performed to ensure accurate background removal and isolate the regions of interest (ROIs) corresponding to the organoids. The segmentation process was crucial in preventing background interference from affecting the extracted features, thereby ensuring that only the organoid structures themselves were analyzed. The segmentation step, integrated into the feature extraction workflow, allowed for precise measurement of both geometric and texture-based features. The dataset of extracted features is available on the aforementioned GitHub repository.

F. MODELS TRAINING AND EVALUATION

Following data pre-processing to ensure high-quality and uniform organoid images, all selected DL/ML models were trained using Jupyter notebooks on Google Colab, leveraging a Tesla V100-SXM2-16GB GPU with 51.0 GB of system RAM. These notebooks are available on the aforementioned GitHub repository.

To evaluate the performance of the different algorithms on the dataset, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values were analyzed. Based on these, classical performance metrics such as Accuracy, Precision, Recall (Sensitivity), Specificity, and F1-score were computed as follows:

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

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

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

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

Moreover, the similarity performance of different classification algorithms was estimated using the MCC (Matthew Correlation Coefficient), also known as Phi Coefficient. This statistical tool, commonly used for model evaluation, provides a value ranging within the interval $[-1, +1]$, and is calculated as follows:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \tag{6}$$

Furthermore, a detailed comparison of the specifications and performance characteristics of six DL models is presented, emphasizing their *size* (measured in megabytes), *complexity* (determined by the number of parameters),

TABLE 3. Key Features extracted for organoid analysis: texture, shape, and color.

Feature	Description
Texture	
contrast	Pixel intensity contrast across the organoid.
correlation	Intensity correlation between adjacent pixels.
dissimilarity	Variation in pixel values.
energy	Texture uniformity from GLCM (Gray-Level Co-occurrence Matrix).
homogeneity	Uniformity in texture.
Shape	
max_solidity	Organoid area to convex hull area ratio.
mean_area	Average organoid area in images.
mean_convex_area	Convex hull area of the organoid.
mean_equivalent_diameter	Circle diameter with same area as organoid.
mean_perimeter	Average boundary length of organoids.
mean_irregularity_index	Deviation from a regular geometric shape.
Color	
mean_intensity	Average pixel brightness within the organoid.
std_intensity	Standard deviation of pixel intensities.

network depth (layer count), and *inference speed* (based on average inference time per image in seconds). All evaluations were conducted on the aforementioned GPU platform.

To estimate the similarity of the different models in the organoid classification, the Jaccard similarity coefficient was calculated for each pair of DL models across all classes of organoids such as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

where A and B are the subsets of organoids identified by the model M_A and M_B in each class, respectively.

IV. EXPERIMENTAL RESULTS

A comparative analysis of six DL and two ML models was performed to classify intestinal organoids. Each model was carefully fine-tuned with a set of optimized training hyperparameters, selected to enhance learning efficiency and model performance. The accuracy of the models was assessed using classical performance metrics. Furthermore, the computational efficiency of each model was examined by measuring runtime and execution time, providing a comprehensive evaluation of both predictive power and practical viability.

A. MODELS CONFIGURATION AND HYPERPARAMETERS TUNING

All DL models were used with pre-trained *ImageNet* weights, excluding the upper layers, and including *2D Global Average Pooling* and a dense layer with 512 units, before a final dense layer corresponding to the four classes with a *softmax* activation function.

The training process employs data augmentation (with horizontal flips, rotation, width and height shifts, zoom and color jitter) to enhance model generalization and reduce overfitting, paired with optimization and *categorical crossentropy* loss, with validation loss monitoring for *early*

stopping, and model checkpointing to preserve the best-performing model.

Table 4 shows the summary of the hyperparameters (with ranges and optimal values highlighted in bold) of all the DLML models exploited in the training step. Such hyperparameters were determined empirically by iteratively fine-tuning values within established ranges, selecting the optimal settings based on performance enhancement. This involved running multiple experiments with different hyperparameter combinations and analyzing the resulting performance metrics described in the next section. Specifically, the impact of hyperparameter values on model performance was assessed by observing changes in key performance metrics. This empirical approach was guided by preliminary experiments that focused on achieving a balance between model performance and computational efficiency. However, while the variations observed were minimal and did not significantly affect the overall results, the ranges of hyperparameter values provide transparency about the tuning process. This ensures reproducibility and highlights the robustness of the models performance across a range of values.

B. MODELS PERFORMANCE EVALUATION

This section provides a comprehensive performance evaluation of the proposed DL/ML models. The comparative analysis is outlined below, highlighting the strengths and limitations of each of them.

Each DL/ML model has been assessed based on the above-mentioned key performance metrics. For each intestinal organoid class, the ability of the DL/ML models to accurately identify images annotated with the correct class was evaluated. Additionally, metrics derived from the Tellu confusion matrix [48] were incorporated into the analysis, providing a meaningful comparison. The inclusion of the Tellu study served as a benchmark for evaluating the proposed DL and ML models against an established method specifically developed for organoid classification tasks.

TABLE 4. Overview of hyperparameter ranges and the optimal values selected for each proposed DL/ML model. The table delineates DL hyperparameters such as learning rate, number of training epochs, type of optimizer, early stopping patience, and batch size, with the optimal values marked in bold. For ML models (SVM and RF), it includes parameters like C, gamma, kernel type, max depth, and min samples split. These configurations were determined through empirical testing to balance model performance and training efficiency.

DL/ML Model	Hyperparameter	Range with optimal
MobileNetV2	*Learning rate*	[0.0001, **0.001**, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, 5, 8, **10**]
	Batch size	[**16**, 32, 64, 128, 256]
DenseNet169	*Learning rate*	[0.0001, **0.001**, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, 5, 8, **10**]
	Batch size	[16, **32**, 64, 128, 256]
ResNet50v2	*Learning rate*	[0.0001, **0.001**, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, 5, 8, **10**]
	Batch size	[16, 32, 64, **128**, 256]
InceptionV3	*Learning rate*	[0.0001, **0.001**, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, 5, 8, **10**]
	Batch size	[16, 32, 64, **128**, 256]
VGG16	*Learning rate*	[0.0001, **0.001**, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, 5, 8, **10**]
	Batch size	[16, 32, 64, **128**, 256]
ViT	*N. of layers*	[16, **24**, 30]
	Learning rate	[**0.0001**, 0.001, 0.01, 0.1]
	Training epochs	[20, **40**, 80, 100, 200]
	Optimizer	[rmsprop, **adam**]
	Early Stopping patience	[2, **5**, 8, 10]
	Batch size	[16, **32**, 64, 128, 256]
SVM	*C*	[0.1, 1, **10**]
	gamma	[scale, auto, **0.1**, 1]
	kernel	[**rbf**, poly, linear, sigmoid]
RF	*max_depth*	[**None**, 10, 20]
	min_samples_split	[2, **5**, 10]

TABLE 5. Comparative metrics of model performance, presenting accuracy, precision, recall, specificity, and F1-score for each tested DL/ML model. The highest values for each metric are highlighted in bold, offering a snapshot of each model's strengths in the organoid classification task.

DL/ML Model	Accuracy	Precision	Recall	Specificity	F1-score	MCC
Tellu [48]	83.63%	79.07%	82.10%	94.04%	80.45%	0.74
MobileNetV2	85.41%	82.51%	83.34%	94.84%	82.62%	0.77
DenseNet169	86.47%	84.35%	81.69%	94.77%	82.71%	0.78
ResNet50v2	79.21%	76.27%	76.11%	92.62%	74.05%	0.68
InceptionV3	86.10%	83.38%	83.92%	94.98%	83.55%	0.78
VGG16	84.19%	81.86%	79.16%	93.91%	79.98%	0.74
ViT	**86.95%**	**87.29%**	**86.95%**	**95.30%**	**87.05%**	**0.80**
RF	64.42%	61.76%	51.15%	84.30%	54.41%	0.41
SVM	68.44%	66.04%	55.90%	86.21%	59.29%	0.48

The following are some observations based on the provided Table 5. Overall, the ViT DL-based model appears to be the most effective across all evaluated metrics, achieving an accuracy of 86.95%, precision of 87.29%, recall of 86.95%, specificity of 95.30%, and an F1-score of 87.05%. This indicates a strong overall performance, marking its

dominance and the most reliable option for organoid image classification tasks. Among the other DL models, InceptionV3 and DenseNet169 also showed strong performances. InceptionV3 achieved an accuracy of 86.10%, precision of 83.38%, recall of 83.92%, and an F1-score of 83.55%, with a high specificity of 94.98%. DenseNet169, on the other hand,

TABLE 6. MCC score performance. The best score is highlighted in bold.

DL/ML Model	Class0	Class1	Class2	Class3
Tellu [48]	0.79	0.63	0.78	0.76
MobileNetV2	0.83	0.66	0.82	0.80
DenseNet169	0.84	0.70	0.83	0.75
ResNet50v2	0.75	0.64	0.70	0.62
InceptionV3	0.82	**0.71**	**0.84**	0.77
VGG16	0.81	0.64	0.80	0.73
ViT	**0.85**	0.70	0.82	**0.83**
SVM	0.55	0.31	0.57	0.50
RF	0.46	0.26	0.46	0.48

demonstrated slightly higher metrics, with an accuracy of 86.47%, precision of 84.35%, and an F1-score of 82.71%. It also maintained a high specificity of 94.77%, indicating its robustness in accurately classifying organoid images. Although Tellu did not outperform the top DL models, it demonstrated a competitive performance, particularly in terms of specificity (94.04%) and F1-score (80.45%). The two ML-based models, RF and SVM, fail to reach in all metrics of the other proposed DL models, where it is possible to notice about 20% less in performance. However, SVM shows superior performance, compared to RF, in all metrics. The lowest value for both is in Recall, 51%-55% where DL models are between 76%-87%, while they reach the highest value in Specificity 84%-86% here instead DL models are between 93%-96%.

The performance of each DL/ML model was evaluated using the MCC score across all classes (see Table 6), where the extreme values of −1 and +1 indicate perfect misclassification and perfect classification, respectively.

Among all models, ViT demonstrated the best overall performance, achieving the highest MCC score in Class0 (0.85) and Class3 (0.83) compared with other models. However, in Class1 and Class2, InceptionV3 outperformed the other models, achieving MCC scores of 0.71 and 0.84, respectively. DenseNet169 also exhibited strong performance, particularly in Class0 with an MCC score of 0.84 and in Class2 with 0.83, positioning it among the top-performing models alongside ViT and InceptionV3. MobileNetV2 demonstrated solid performance across several classes, notably in Class0 with an MCC score of 0.83 and Class3 with 0.80, showing its capability despite being a more lightweight model. Tellu also performed well, particularly in Class0 with an MCC score of 0.79 and Class2 with a score of 0.78, highlighting the strengths of DL models over traditional ML models like SVM and RF, which demonstrated limitations in all classes, particularly in Class1 and Class3.

This underscores the power of DL in handling complex image data. Hence also the motivation to base this work mainly on the DL approaches, as they consistently demonstrated superior performance in classifying organoid images, even in the presence of high inter-class variability and subtle morphological differences. The ability of DL models to automatically learn and extract relevant features from raw data, without requiring manual feature engineering, is a significant advantage over traditional ML methods.

Furthermore, the robustness of DL models across various performance metrics highlights their adaptability to different data distributions and imaging conditions.

However, the choice of DL model could depend on the specific use case. It is also noteworthy that the differences in performance are relatively modest across DL models, which indicates that each has its merits and could be useful in different scenarios or with different optimization and tuning efforts. The flexibility and scalability of DL models, particularly when dealing with large and complex datasets, further justify the focus of this work on these models in organoid image classification.

The performance of the DL models was then evaluated for each class, providing detailed insights into their ability to accurately classify images across the different four classes.

All TP and FP statistics are represented in Figure 3 with a stacked barplot showing correct identification for each class performed using proposed DL models. The evaluation of model performance revealed the following insights:

Class0. DenseNet169 demonstrated the highest performance with 93.82% true positives, followed closely by VGG16 and ViT.

Class1. InceptionV3 showed superior performance with 83.21% correct images identification, followed by ResNet50v2 and ViT.

Class2. MobileNetV2 shows a better performance with 94.26% images correctly identified followed by VGG16 and DenseNet169.

Class3. ResNet50v2 identified 82.67% of images correctly followed by ViT and InceptionV3.

Figure 4 shows the results in a series of confusion matrices for the six DL models by describing the performance of the DL classification model on all classes. A detailed explanation is given below:

Class0. DenseNet169 is the model accomplishing the best performance to correctly identify 1.214 (93.82%) Class0 images, followed by VGG16 with 1.202 (92.89%), ViT 1.178 (91.04%), MobileNetV2 1.161 (89.72%), InceptionV3 1.152 (89.03%) and ResNet50v2 1.106 (85.47%). All models share a set of FP generally with Class1, with a range between 3.5% (ResNet50v2) and 8.1% (MobileNetV2), and the other FPs belong to Class3 with a range between 1.47% (DenseNet169) and 10.9% (ResNet50v2).

Class1. InceptionV3 shows a 456 (83.21%) correctly annotated, followed by ResNet50v2 434 (79.20%), ViT 430 (78.47%), DenseNet169 402 (73.36%), MobileNetV2 395 (72.08%) and VGG16 361 (65.88%). To be noted that all models show an over-classification for Class2 (21.1% - 8.76%) with respect Class0 (13.14% - 6.57%), only ResNet50v2 have an opposite trend assigned the images in Class0 for 14.9% and the rest on Class2 2.19%.

Class2. MobileNetV2 model performed an accurate classification with 378 (94.26%) correctly images identified, followed by VGG16 364 (90.77%), DenseNet169 362 (90.27%), ViT 357 (89.03%), InceptionV3 340 (84.79%) and ResNet50v2 229 (57.11%). Generally, the FPs are annotated
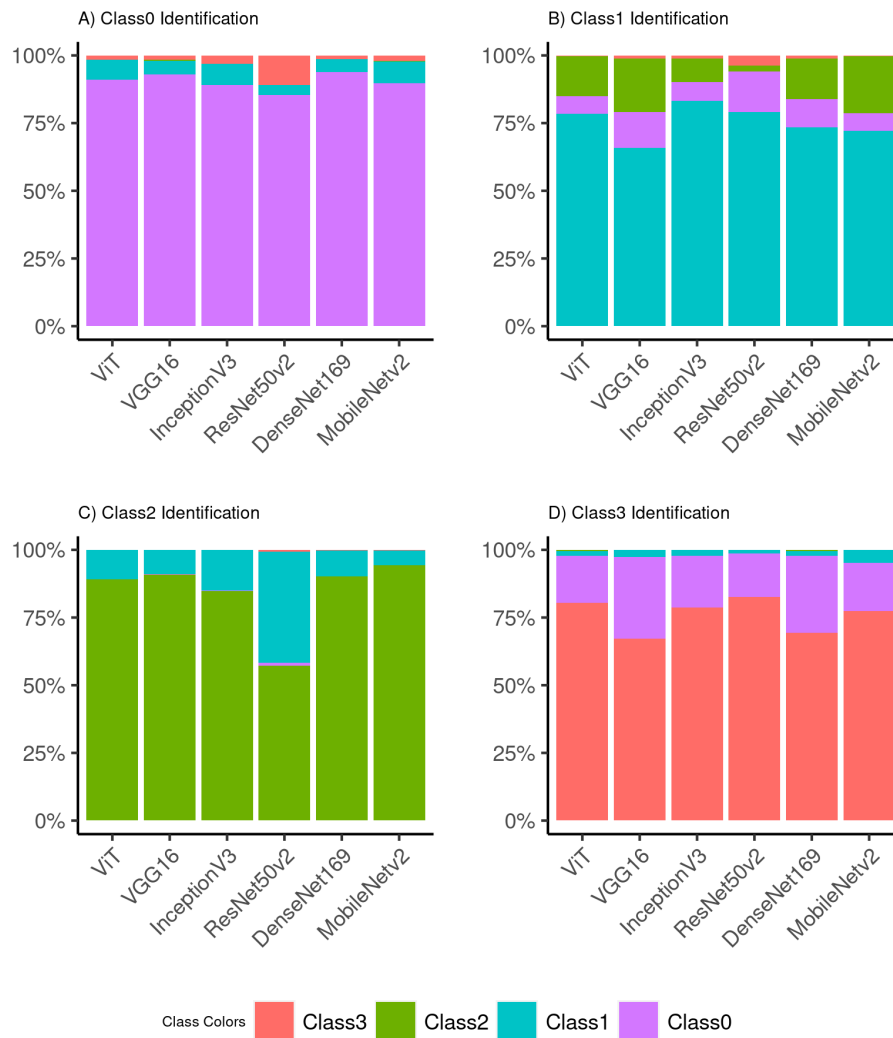
FIGURE 3. Stacked barplot showing correct identification distribution for each class across the proposed DL models. Each subplot (A, B, C, D) corresponds to a specific class (Class0, Class1, Class2, Class3). The colors represent the different classes as identified by each DL model.

in Class1 (MobileNetV2 5.49% - ResNet50v2 40.90%), the immediately preceding stage.

Class3. ResNet50v2 model has better accuracy, identifying 186 Class3 (82.67%), followed by ViT with 181 (80.44%), InceptionV3 with 177 (78.67%), MobileNetV2 with 174 (77.33%), DenseNet169 with 156 (69.33%) and VGG16 151 (67.11%).

Figure 5 shows a comparative analysis of the Receiver Operating Characteristic (ROC) curves for the four distinct classes as modeled by six DL architectures, where each panel—labeled from (A) to (F) represents the model's capability to distinguish among the classes, with the curves for Classes 0 through 3 illustrated distinctly. The plots illustrate the diagnostic ability of a classifier system as its discrimination threshold is varied. The ROC curves demonstrate the true positive rate against the false positive rate for each class, providing insights into the DL models'

classification capabilities. Notably, the area under the curve (AUC) offers a quantifiable measure of the model's ability to differentiate between classes. This comprehensive evaluation across multiple classes and models facilitated a nuanced understanding of each model's strengths and limitations in a multi-class classification context. Overall, the ROC curves suggest that all models have strong discriminative abilities for the classes they are tested on. DenseNet169 and ViT models show particularly high AUC values, suggesting that these models have superior performance in classifying the four classes. The high AUC values for the *Class2* across all models indicate that this class is the easiest to classify, whereas other classes may present more of a challenge, as shown by the slightly lower AUC values of *Class1*.

The pairwise similarity of predictions across different DL models was assessed by calculating the Jaccard coefficient
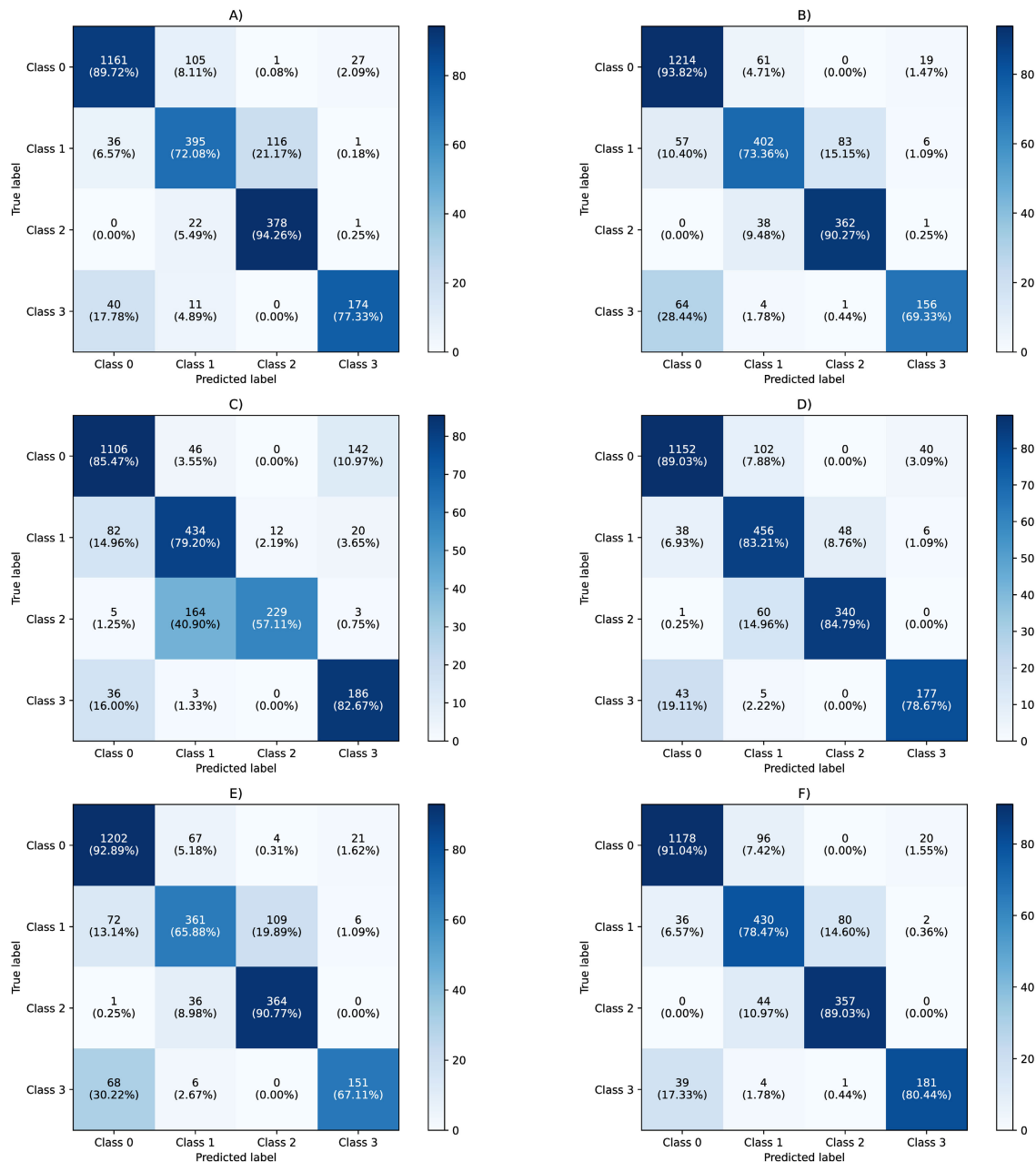
FIGURE 4. Confusion matrices showcasing the performance of the six DL models in the organoid multiclass classification task. Each matrix corresponds to A) MobileNetV2, B) DenseNet169, C) ResNet50v2, D) InceptionV3, E) VGG16, and F) ViT, with organoid counts and percentages indicating true labels versus predicted labels.

TABLE 7. Specifications and performance of the proposed DL models. The table reports the model size (in MB), model complexity in terms of millions of parameters (M), network depth as measured by the number of layers, training time per epoch (in seconds), and inference time per image (in seconds).

DL Model	Size (MB)	Complexity (M)	Network depth	Training time (sec/epoch)	Inference time (sec/image)
Tellu [48]	14.3	7.2	213	*N/A**	0.0198
ViT	343.0	86	24	1100.0	0.0437
VGG16	183.1	138.4	16	317.0	0.0111
InceptionV3	275.5	23.9	189	279.0	0.0105
ResNet50v2	295.9	25.6	103	285.0	0.0073
DenseNet169	163.5	14.3	338	340.0	0.0196
MobileNetV2	35.6	3.5	105	279.0	0.0063

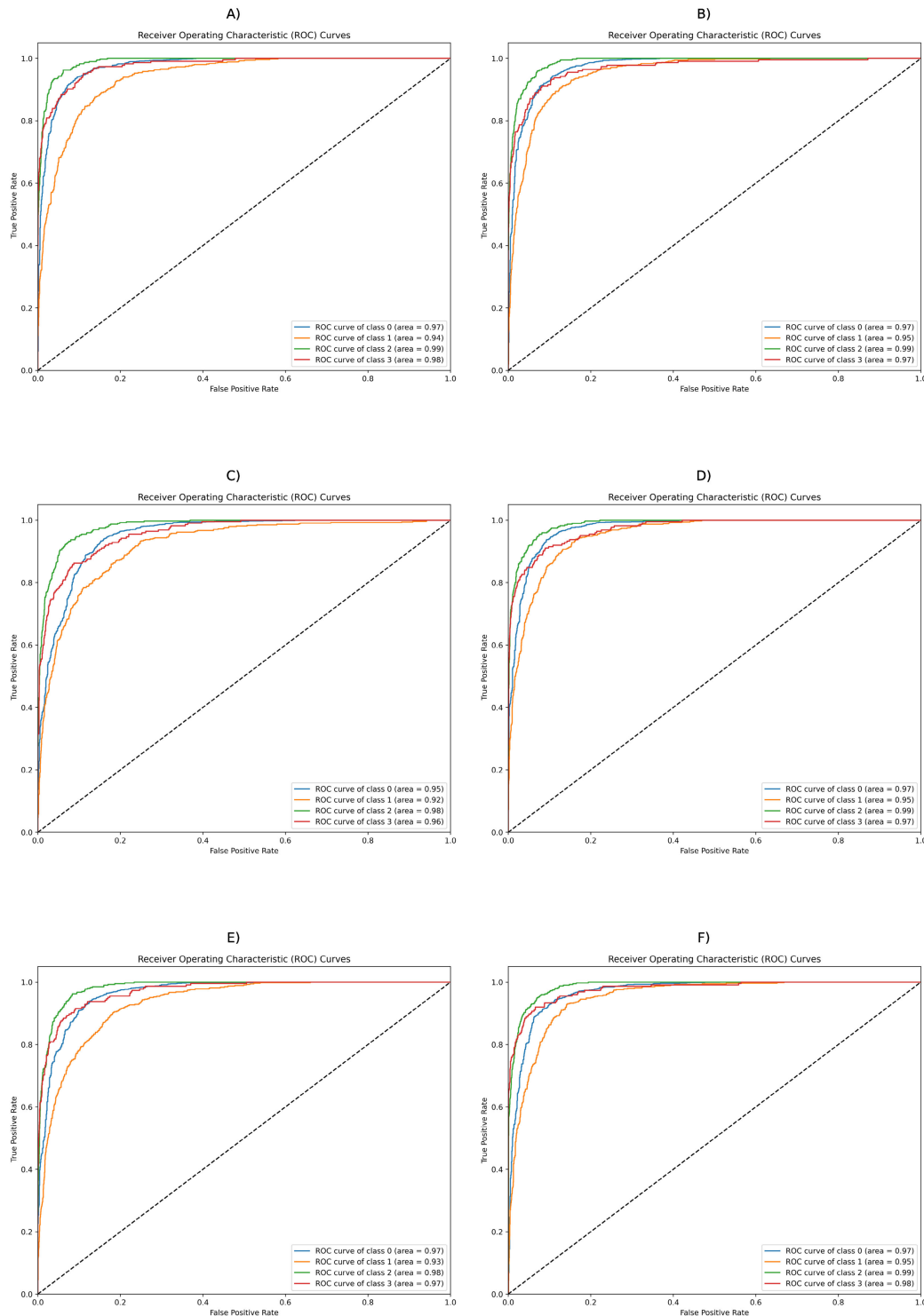*Training time per epoch is not reported in the original source.

FIGURE 5. ROC curves comparison for the six DL models across four classes, utilizing the area under the curve (AUC) metric to indicate classification accuracy in the classification task. Each subplot reflects the true positive rate versus the false positive rate for A) MobileNetV2, B DenseNet169, C) ResNet50v2, D) InceptionV3, E) VGG16, and F) ViT.

for each class and each pair of DL models (refer to Figure 6). This analysis showed that the models generally exhibited comparable performance. In particular, as in Class0, the DL models have comparable performance, with

coefficients varying between 0.80 and 0.91. Notably, for Class1, the DenseNet169-VGG16 pair showed the lowest similarity score of 0.52, whereas the ViT-InceptionV3 pair achieved the highest score of 0.79. Similarly, for Class 2,
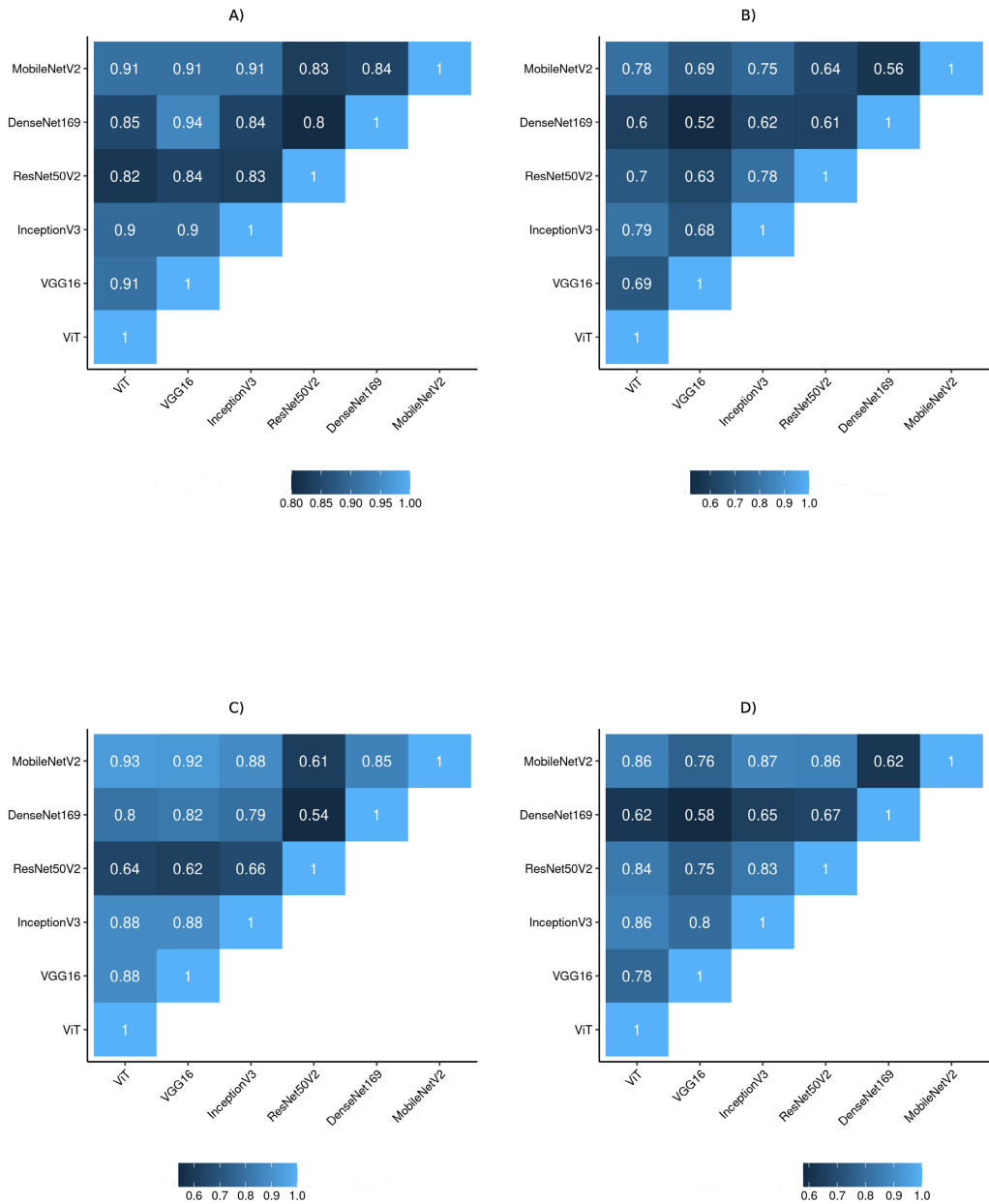
FIGURE 6. Jaccard similarity matrices for different DL models. The matrices show the similarity coefficients between each pair of DL models across different organoid classes, from Class0 (A) to Class3 (D). The color intensity indicates the level of similarity, with lighter shades representing higher similarity scores.

MobileNetV2-ViT scored highest at 0.93, demonstrating better alignment in their classification outcomes compared to other model pairs. Finally, for Class3, InceptionV3-MobileNetV2 achieves a score of 0.87, while the least similar pair is VGG16-DenseNet169 with 0.58.

C. MODELS EFFICIENCY

Given that the choice of an appropriate classification model must also consider task-specific requirements, infrastructural constraints, and the trade-off between performance metrics, however, it is important to consider in addition to classic classification metrics, other factors such as computational cost, inference time, and model complexity.

Table 7 provides an overview of the specifications and performance characteristics of the proposed DL models, highlighting their efficiency in the classification of organoids. The table includes key information such as model size (in MB), complexity (measured in millions of parameters), network depth (number of layers), training time per epoch (in seconds), and inference time per image (in seconds).

Although the ViT model represents the DL model that best generalizes the classification task, it is also the largest model in terms of file size, at 343.0 MB, and has a moderately high complexity, with 86 million parameters. Despite the smaller number of layers (24), it has a longer inference time of 0.0437 seconds, probably due to the

computational intensity of the transform architecture. This could make ViT less suitable for real-time applications, but potentially more powerful in terms of learning capabilities. On the other side, with 35.6 MB, MobileNetV2 is the smallest model with the least complexity, housing only 3.5 million parameters. With 105 layers, it has a moderate depth and the first-fastest inference time at 0.0063 seconds, making it an ideal candidate for mobile or edge devices. VGG16 is considerably large at 183.1 MB and has a high complexity with 138.4 million parameters. It has a shallower network depth of 16 layers and a moderate inference time of 0.0111 seconds. In addition, the model size and the number of parameters of VGG16 indicate a significant memory overhead, but the inference speed is a reasonable one. InceptionV3 is a large-sized model at 275.5 MB with a moderate complexity level of 23.9 million parameters. This model is notably deep with 189 layers, which is reflected in its longer inference time of 0.0105 seconds compared to some shallower models. ResNet50v2 model is relatively more complex than InceptionV3, with a size of 295.9 MB and 25.6 million parameters. It has a significant network depth of 103 layers but boasts the fastest inference time among the models at 0.0073 seconds. ResNet50v2's architecture optimizes the inference efficiency, making it a compelling choice for applications requiring rapid results. DenseNet169 is on the smaller side in terms of size at 163.5 MB and has the lowest complexity with 14.3 million parameters. It has a high network depth of 338 layers, yet maintains a relatively quick inference time of 0.0196 seconds. Its low complexity-to-depth ratio and quick inference time make it an attractive model for environments with some computational constraints.

When compared with Tellu, it stands out due to its smaller size of 14.3 MB, making it lightweight and less demanding in terms of storage. Despite its modest complexity with 7.2 million parameters and a network depth of 213 layers, Tellu manages to maintain a reasonable inference time of 0.0198 seconds. This efficiency, combined with its compact size, suggests that Tellu is a strong candidate for applications in resource-limited environments.

Figure 7 helps to quickly assess the trade-offs between speed and performance of different proposed DL models. It shows the trade-off between inference time (run-time in seconds) on the abscissae and model accuracy (accuracy in percentage) on the ordinates, with the size of the bubbles representing the complexity of the model (in terms of the number of parameters and the weight of the model file). Larger and more complex DL models, such as ViT, show higher accuracy but also longer inference time. In contrast, smaller and less complex DL models such as MobileNetV2 exhibit shorter inference times at the expense of slightly reduced accuracy. Overall, these specifications reveal an inherent balance between a model's size and intricacy against its speed of inference, underscoring the necessity of selecting a DL model that aligns with the unique demands and limitations of the intended application environment.

V. DISCUSSION

This study presents a systematic comparative analysis of DL models for the accurate classification of all organoid classes. Specifically, five CNN-based DL models and the recent ViT were selected to serve as a guide for researchers seeking the most appropriate model for the automatic classification of intestinal organoids. For this purpose, a publicly available dataset, previously created and annotated, consisting of 840 images of intestinal organoids, was utilized. Using a Python script, a bounding box of 224×224 pixels was cropped to annotate each image. This image scaling step ensured that all models received input data of uniform dimensions, thereby reducing variability and maintaining compatibility with the pre-trained DL architectures. The final dataset, containing 23,063 annotations, was divided into three subsets: training (70%), validation (20%), and testing (10%), ensuring balanced class representation in each subset. The performance of the DL models (as well as widely used ML models for comparison) was evaluated based on two key criteria: (i) accurate identification of different intestinal organoid classes (Organoid0, Organoid1, Organoid3, Spheroid) and (ii) efficiency, assessed through runtime and execution time. Performance metrics, including Accuracy, Precision, Recall, Specificity, F-measure, and MCC, were employed to assess the DL and ML models. The results clearly demonstrated the superiority of DL models compared to traditional ML approaches, highlighting the importance of leveraging DL methodologies in complex image classification tasks, such as those involving organoids.

A. MODELS PERFORMANCE INSIGHTS

ViT-based model demonstrates superior general performance across the above-cited proposed metrics. However, for readers seeking a balance between accuracy and computational efficiency, InceptionV3 emerges as an optimal compromise, offering strong performance metrics while maintaining reasonable inference times.

In the identification of Class0, DenseNet169 shows the best performance compared to other models. Most of the remaining annotations were misclassified as Class1, with Class3 as the second instance. Interestingly, the ResNet50v2 model assigned fewer images to Class1 but incorrectly classified these images into Class3. For Class1, the InceptionV3 model reaches the best performances, with incorrectly annotated images distributed equally between Class0 and Class2. All other models predominantly misclassified these images into Class2, whereas the ResNet50v2 model, notably, labeled most of them as Class0. For Class2, MobileNetV2 shows better performance compared to other models, with the remaining misclassified images predominantly labeled as Class1. Finally, in the case of Class3, ResNet50v2 achieves the highest rate of correct annotations, with most misclassified images assigned to Class0, followed by Class1 with a lower frequency.

In the overall results, the ViT model emerges as the superior choice for intestinal organoid classification as
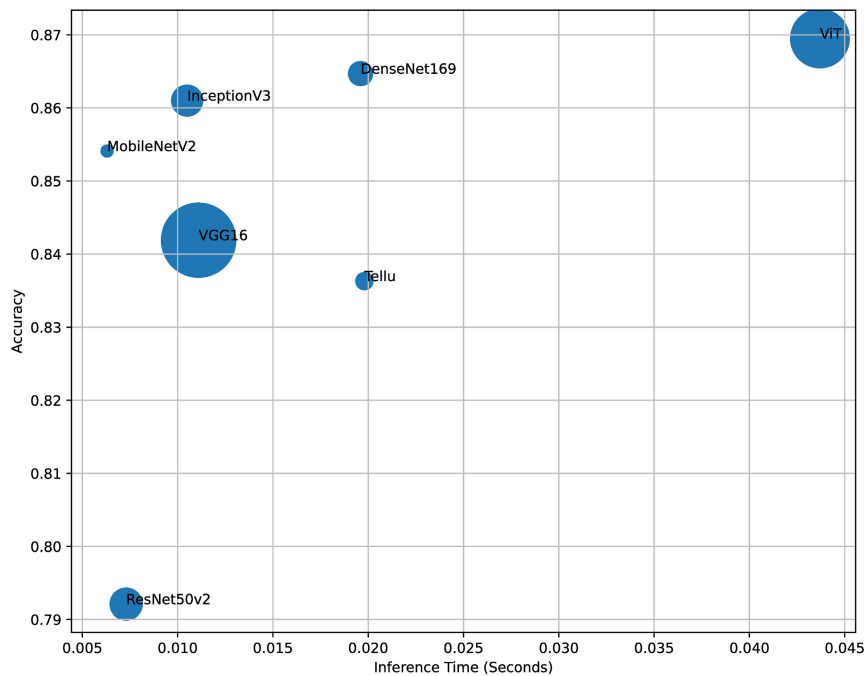
FIGURE 7. Trade-off between DL model complexity, run-time and accuracy. The x-axis represents inference time in seconds; The y-axis shows the accuracy of each model, measured as a percentage; The size of the bubbles corresponds to the complexity of the model, where larger bubbles suggest more complex models.

evidenced by the metrics reported in Table 5 and MCC score presented in Table 6. Additionally, efficiency was evaluated by comparing runtime and execution time, which are critical parameters for real-time applications. As illustrated in Figure 7, a clear trade-off exists between the size and complexity of a model and its inference time. Larger and more complex models, such as ViT, demonstrate greater feature extraction capabilities and superior performance but require higher computational resources. Conversely, lightweight models like MobileNetV2 are optimized for efficiency, offering significant advantages in speed and size, making them particularly suitable for resource-constrained environments or applications requiring rapid decision-making. Based on these insights, InceptionV3 appears to be a viable option for future organoid classification tasks, balancing strong performance with computational efficiency.

B. PITFALLS OF CLASSIFICATION MODELS

This section provides an in-depth analysis of the cases in which each model misclassifies images, aiming to better understand the sources of error. Certain morphological features, such as the compactness of organoids and irregularities in shape, which are less apparent in 2D images, were identified as factors contributing to misclassification. Additionally, variations in image properties, including brightness and contrast, were found to exacerbate errors, particularly in images captured under suboptimal illumination conditions.

The misclassification between Class0 (Cyst) and Class3 (Spheroid) was particularly notable. These two classes share

overlapping morphological features, such as similar circular shapes and wall thickness, which challenge the models' ability to distinguish them effectively. Also, suboptimal lighting and inconsistent contrast amplified these challenges, reducing the model's overall performance in distinguishing these classes.

Based on the experimental results, image acquisition practices are recommended to improve quality and classification. Particularly, the proper use of contrast is essential, as images with low contrast were found to be more challenging to classify accurately. Additionally, excessive variability in illumination conditions during image acquisition should be avoided. Most importantly, the establishment of a standardized imaging protocol is highly recommended to reduce inter-sampling variability. Among the DL models, those with more sophisticated feature extraction capabilities, such as ViT, demonstrated a slightly better ability to handle variations in image contrast and brightness. This observation suggests that focusing on model architectures inherently more sensitive to such variations could provide an avenue for improving overall accuracy. Furthermore, it was noted that misclassified images were often assigned to a proximal class. This may be attributed to shared features between certain classes, as illustrated in Figure 2. For instance, images from Class0 and Class3 shared more morphological features, such as border and shape, despite being temporally antipodal. Instead, greater diversity was observed between Class1 and Class2, making them more distinguishable.

To provide further transparency, an Excel file containing the true labels of each image and the predictions made by

each model has been made available in the GitHub repository. This file highlights that the majority of misclassified images are shared across multiple models.

Upon analyzing the results, it was found that 18 images were consistently misclassified by all DL models. Of these, 8 belonged to Class1, 1 to Class2, and 9 to Class3. Specifically, for Class1, most models assigned the images to either Class0 or Class2. For the single misclassified Class2 image, all models misclassified it as Class1, while ResNet50v2 incorrectly classified it as Class0. Similarly, for the misclassified Class3 images, all DL models assigned them to Class0.

Figure 8 presents a set of representative images highlighting different misclassification scenarios. The first column (Panel A) illustrates cases where ViT correctly classified organoid images that all other DL/ML models misclassified. These images exhibit complex morphological features that likely benefit from ViT's self-attention mechanism, which better captures long-range dependencies and structural cues across the image. Conversely, Panel B displays images consistently misclassified by all models, indicating that morphological similarities between classes, such as the overlap between Class0 (Cyst) and Class3 (Spheroid), pose significant challenges for automated classification. These misclassifications highlight the inherent limitations of DL models in distinguishing classes with overlapping features, such as circular shapes and similar wall thicknesses. Finally, Panel C includes the few cases where proposed conventional ML models (SVM and RF) outperformed DL-based models, likely due to their reliance on handcrafted features rather than learned spatial representations. The ability of ML models to leverage these explicit descriptors contributed to their improved performance in classifying organoids with subtle but critical morphological variations.

C. RESEARCH LIMITATIONS

Despite its numerous advantages, this research also has limitations: i) One of the key challenges encountered in this study was the issue of class imbalance. The imbalance in the dataset may still have influenced the model's ability to generalize effectively across all classes. However, this could lead to a bias where the model performs better on the majority classes at the expense of minority classes. ii) DL models utilized in this work, although powerful, come with significant computational costs. Training these models required substantial computational resources, which could limit their scalability, especially in settings with limited access to high-performance computing infrastructure. This complexity also raised concerns about the feasibility of deploying these models in real-time or resource-constrained environments. iii) Another limitation was the interpretability of the DL models compared to traditional ML approaches. While DL models automatically learn features during the training process, this "black-box" nature makes it challenging to understand the underlying factors driving the model's predictions. In contrast, traditional ML models allow for manual feature extraction, providing clearer insights into the
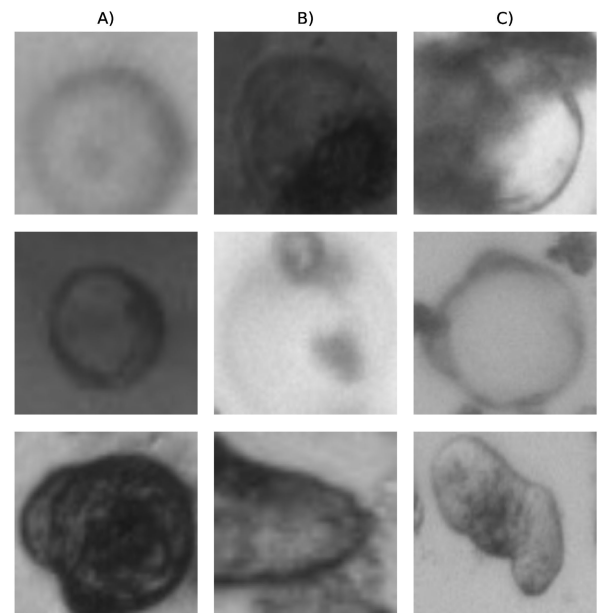

FIGURE 8. Representative cases of misclassification. Examples of organoid images correctly classified exclusively by the ViT model (A), misclassified by all models (B), and correctly classified exclusively by conventional ML models (C). In each panel, the first two images upward correspond to Class3, while the last image at the bottom refers to Class1.

decision-making process. To address this limitation, future work will explore the adoption of Explainable AI (XAI) techniques to enhance the interpretability of the proposed DL models. These methods will aim to highlight the key features contributing to each classification decision, offering a more transparent understanding of the model's inner workings and fostering greater trust in its predictions. iv) Moreover, while DenseNet169, MobileNetV2, ResNet50v2, InceptionV3, VGG16, and ViT are commonly used for 2D brightfield organoid images, they can be adapted or extended to handle volumetric 3D images. As 3D analysis becomes more common in organoid and tissue studies, there's a growing need to apply DL models to analyze such data. Adapting these models for 3D volumetric images requires careful consideration of architectural modifications, data preprocessing steps, and computational resources due to the increased complexity of 3D data compared to 2D images. However, with appropriate adjustments, these models can be valuable tools for analyzing volumetric organoid and tissue images.

VI. CONCLUSION AND FUTURE WORKS

This study provides a comprehensive investigation into the main principles of DL models for organoid classification tasks. Advancements in the classification tools and models of organoids represent a pivotal asset for researchers. To the best of available knowledge, this work constitutes the first comprehensive benchmark assessing the robustness of several DL models in the automatic classification of intestinal organoids. A broad overview of the most popular DL models was presented, with the aim of assisting readers in the choice of

the best model for accurate organoid classification, while also providing a detailed understanding of the current research landscape in this field. A series of extensive experiments were conducted to benchmark the robustness of organoid image classifiers, specifically designed to standardize DL model evaluation, enhance phenotypic profiling reliability, and accelerate drug discovery. Unlike previous studies such as ''Tellu'', which primarily focused on object detection and classification, this work introduces an innovative approach by systematically benchmarking multiple DL architectures, including CNNs and ViT, rather than relying on a single model (YOLO-based).

Looking ahead, the integration of these DL models as plug-ins within image analysis software could revolutionize the field. This would not only streamline the organoid classification process but also facilitate real-time image identification, thereby enhancing both research efficiency and analytical precision. Moreover, the practical implications of these findings extend beyond classification accuracy. Embedding these models, particularly ViT, into clinical workflows could enhance personalized medicine by providing rapid and precise analysis of patient-derived organoids. In a typical workflow, classified morphological phenotypes could be linked to patient-specific drug response profiles, supporting tailored treatment strategies. Additionally, these models could be integrated into high-throughput drug screening pipelines to automatically evaluate large-scale image datasets, enabling the identification of subtle morphological changes induced by candidate compounds and accelerating therapeutic development.

Based on the misclassification problems observed with Class0 and Class3 organoids, future work will focus on enhancing model architectures, implementing new data augmentation and preprocessing techniques, and utilizing advanced feature extraction methods. The latter will involve texture analysis, shape descriptors, or edge detection to better differentiate classes. Additionally, efforts will be made to balance the dataset and apply interpretability tools and XAI techniques, such as Grad-CAM and SHAP, to enhance the interpretability and clinical applicability of DL-based organoid classification. Specifically, attention-based visualization methods (e.g., attention rollout maps for ViT) will be explored to highlight relevant morphological features across organoid classes. Moreover, future research will focus on expanding the diversity of training datasets and exploring the applicability of these DL models to other forms of biological image analysis, thereby enhancing their utility and robustness in varied research contexts.

REFERENCES

[1] D. M. Parkin, P. Pisani, and J. Ferlay, "Global cancer statistics," *CA, A Cancer J. Clinicians*, vol. 49, no. 1, pp. 33–64, Jan. 1999.

[2] Y. Xi and P. Xu, "Global colorectal cancer burden in 2020 and projections to 2040," *Translational Oncol.*, vol. 14, no. 10, Oct. 2021, Art. no. 101174.

[3] L. Ricci-Vitiani, D. G. Lombardi, E. Pilozzi, M. Biffoni, M. Todaro, C. Peschle, and R. De Maria, "Identification and expansion of human colon-cancer-initiating cells," *Nature*, vol. 445, no. 7123, pp. 111–115, Jan. 2007.

[4] P. Dalerba, S. J. Dylla, I.-K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. Shelton, G. Parmiani, C. Castelli, and M. F. Clarke, "Phenotypic characterization of human colorectal cancer stem cells," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 24, pp. 10158–10163, Jun. 2007.

[5] Y. Huang, Z. Huang, Z. Tang, Y. Chen, M. Huang, H. Liu, W. Huang, Q. Ye, and B. Jia, "Research progress, challenges, and breakthroughs of organoids as disease models," *Frontiers Cell Develop. Biol.*, vol. 9, Nov. 2021, Art. no. 740574.

[6] Z. Zhao, X. Chen, A. M. Dowbaj, A. Sljukic, K. Bratlie, L. Lin, E. L. S. Fong, G. M. Balachander, Z. Chen, A. Soragni, M. Huch, Y. A. Zeng, Q. Wang, and H. Yu, "Organoids," *Nature Rev. Methods Primers*, vol. 2, no. 1, p. 94, 2022.

[7] T. Sato, R. G. Vries, H. J. Snippert, M. van de Wetering, N. Barker, D. E. Stange, J. H. van Es, A. Abo, P. Kujala, P. J. Peters, and H. Clevers, "Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche," *Nature*, vol. 459, no. 7244, pp. 262–265, May 2009.

[8] T. Sato, D. E. Stange, M. Ferrante, R. G. J. Vries, J. H. van Es, S. van den Brink, W. J. van Houdt, A. Pronk, J. van Gorp, P. D. Siersema, and H. Clevers, "Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium," *Gastroenterology*, vol. 141, no. 5, pp. 1762–1772, Nov. 2011.

[9] J. Qu, F. S. Kalyani, L. Liu, T. Cheng, and L. Chen, "Tumor organoids: Synergistic applications, current challenges, and future prospects in cancer therapy," *Cancer Commun.*, vol. 41, no. 12, pp. 1331–1353, Dec. 2021.

[10] S. Mao, Y. Pang, T. Liu, Y. Shao, J. He, H. Yang, Y. Mao, and W. Sun, "Bioprinting of in vitro tumor models for personalized cancer treatment: A review," *Biofabrication*, vol. 12, no. 4, Jul. 2020, Art. no. 042001.

[11] B. L. LeSavage, R. A. Suhar, N. Broguiere, M. P. Lutolf, and S. C. Heilshorn, "Next-generation cancer organoids," *Nature Mater.*, vol. 21, no. 2, pp. 143–159, Feb. 2022.

[12] Z. Fang, P. Li, F. Du, L. Shang, and L. Li, "The role of organoids in cancer research," *Exp. Hematol. Oncol.*, vol. 12, no. 1, p. 69, Aug. 2023.

[13] J. Drost and H. Clevers, "Organoids in cancer research," *Nature Rev. Cancer*, vol. 18, no. 7, pp. 407–418, Jul. 2018.

[14] L. Yang, B. Liu, H. Chen, R. Gao, K. Huang, Q. Guo, F. Li, W. Chen, and J. He, "Progress in the application of organoids to breast cancer research," *J. Cellular Mol. Med.*, vol. 24, no. 10, pp. 5420–5427, May 2020.

[15] J. A. Brassard and M. P. Lutolf, "Engineering stem cell self-organization to build better organoids," *Cell Stem Cell*, vol. 24, no. 6, pp. 860–876, Jun. 2019.

[16] D. Serra, U. Mayr, A. Boni, I. Lukonin, M. Rempfler, L. Challet Meylan, M. B. Stadler, P. Strnad, P. Papasaikas, D. Vischi, A. Waldt, G. Roma, and P. Liberali, "Self-organization and symmetry breaking in intestinal organoid development," *Nature*, vol. 569, no. 7754, pp. 66–72, May 2019.

[17] P. Andrei, P. Battuello, G. Grasso, E. Rovera, N. Tesio, and A. Bardelli, "Integrated approaches for precision oncology in colorectal cancer: The more you know, the better," *Seminars Cancer Biol.*, vol. 84, pp. 199–213, Sep. 2022.

[18] S. Montes-Olivas, D. Legge, A. Lund, A. G. Fletcher, A. C. Williams, L. Marucci, and M. Homer, "In-silico and in-vitro morphometric analysis of intestinal organoids," *PLOS Comput. Biol.*, vol. 19, no. 8, Aug. 2023, Art. no. e1011386.

[19] J. Ko, S. Hyung, S. Cheong, Y. Chung, and N. Li Jeon, "Revealing the clinical potential of high-resolution organoids," *Adv. Drug Del. Rev.*, vol. 207, Apr. 2024, Art. no. 115202.

[20] L. Smirnova and T. Hartung, "The promise and potential of brain organoids," *Adv. Healthcare Mater.*, vol. 13, no. 21, Aug. 2024, Art. no. 2302745.

[21] M. P. Viana et al., "Integrated intracellular organization and its variations in human iPS cells," *Nature*, vol. 613, no. 7943, pp. 345–354, Jan. 2023.

[22] J. H. Levine, E. F. Simonds, S. C. Bendall, K. L. Davis, E.-A.-D. Amir, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, R. Finck, A. L. Gedman, I. Radtke, J. R. Downing, D. Pe'er, and G. P. Nolan, "Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis," *Cell*, vol. 162, no. 1, pp. 184–197, Jul. 2015.

[23] P. Stüve, B. Nerb, S. Harrer, M. Wuttke, M. Feuerer, H. Junger, E. Eggenhofer, B. Lungu, S. Laslau, and U. Ritter, "Analysis of organoid and immune cell co-cultures by machine learning-empowered image cytometry," *Frontiers Med.*, vol. 10, Jan. 2024, Art. no. 1274482.

[24] C. Sommer and D. W. Gerlich, "Machine learning in cell biology–teaching computers to recognize phenotypes," *J. Cell Sci.*, vol. 126, no. 24, pp. 5529–5539, Jan. 2013.

[25] B. T. Grys, D. S. Lo, N. Sahin, O. Z. Kraus, Q. Morris, C. Boone, and B. J. Andrews, "Machine learning and computer vision approaches for phenotypic profiling," *J. Cell Biol.*, vol. 216, no. 1, pp. 65–71, Jan. 2017.

[26] C. Scheeder, F. Heigwer, and M. Boutros, "Machine learning and image-based profiling in drug discovery," *Current Opinion Syst. Biol.*, vol. 10, pp. 43–52, Aug. 2018.

[27] I. Lukonin, D. Serra, L. Meylan, K. Volkmann, J. Baaten, R. Zhao, S. Meeusen, K. Colman, F. Maurer, M. B. Stadler, J. Jenkins, and P. Liberali, "Phenotypic landscape of intestinal organoid regeneration," *Nature*, vol. 586, no. 7828, pp. 275–280, Oct. 2020.

[28] T. Kassis, V. Hernandez-Gordillo, R. Langer, and L. G. Griffith, "OrgaQuant: Human intestinal organoid localization and quantification using deep convolutional neural networks," *Sci. Rep.*, vol. 9, no. 1, p. 12479, Aug. 2019.

[29] J. Matthews, B. Schuster, S. S. Kashaf, P. Liu, M. Bilgic, A. Rzhetsky, and S. Tay, "Organoid: A versatile deep learning platform for organoid image analysis," *bioRxiv*, vol. 2022, pp. 1–18, Jan. 2022.

[30] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, Jul. 2012.

[31] C.-L. Chiu and N. Clack, "Napari: A Python multi-dimensional image viewer platform for the research community," *Microsc. Microanalysis*, vol. 28, no. S1, pp. 1576–1577, Aug. 2022.

[32] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, J. A. James, M. Salto-Tellez, and P. W. Hamilton, "QuPath: Open source software for digital pathology image analysis," *Sci. Rep.*, vol. 7, no. 1, pp. 1–7, Dec. 2017.

[33] N. Gritti, J. L. Lim, K. Anlas, M. Pandya, G. Aalderink, G. Martínez-Ara, and V. Trivedi, "MOrgAna: Accessible quantitative analysis of organoids with machine learning," *Development*, vol. 148, no. 18, Sep. 2021, Art. no. 199611.

[34] R. N. U. Kok, L. Hebert, G. Huelsz-Prince, Y. J. Goos, X. Zheng, K. Bozek, G. J. Stephens, S. J. Tans, and J. S. van Zon, "OrganoidTracker: Efficient cell tracking using machine learning and manual error correction," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0240802.

[35] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, and A. Kreshuk, "Ilastik: Interactive machine learning for (bio) image analysis," *Nature methods*, vol. 16, no. 12, pp. 1226–1232, 2019.

[36] D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, and A. Goodman, "CellProfiler 4: Improvements in speed, utility and usability," *BMC Bioinf.*, vol. 22, no. 1, pp. 1–11, Dec. 2021.

[37] V. Zinchenko, J. Hugger, V. Uhlmann, D. Arendt, and A. Kreshuk, "MorphoFeatures for unsupervised exploration of cell types, tissues, and organs in volume electron microscopy," *eLife*, vol. 12, p. 80918, Feb. 2023, doi: 10.7554/elife.80918.

[38] M. De Vries, L. Dent, N. Curry, L. Rowe-Brown, A. Tyson, C. Dunsby, and C. Bakal, "3D single-cell shape analysis of cancer cells using geometric deep learning," *bioRxiv*, vol. 2022, pp. 1–23, Jun. 2022.

[39] J. Betge et al., "The drug-induced phenotypic landscape of colorectal cancer organoids," *Nature Commun.*, vol. 13, no. 1, p. 3135, Jun. 2022.

[40] X. Bian, G. Li, C. Wang, W. Liu, X. Lin, Z. Chen, M. Cheung, and X. Luo, "A deep learning model for detection and tracking in high-throughput images of organoid," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104490.

[41] E. Kegeles, A. Naumov, E. A. Karpulevich, P. Volchkov, and P. Baranov, "Convolutional neural networks can predict retinal differentiation in retinal organoids," *Frontiers Cellular Neurosci.*, vol. 14, p. 171, Jul. 2020.

[42] X. Du, Z. Chen, Q. Li, S. Yang, L. Jiang, Y. Yang, Y. Li, and Z. Gu, "Organoids revealed: Morphological analysis of the profound next generation in-vitro model with artificial intelligence," *Bio-Design Manuf.*, vol. 6, no. 3, pp. 319–339, May 2023.

[43] N. Zhao, R. T. Powell, X. Yuan, G. Bae, K. P. Roarty, F. Stossi, M. Strempfl, M. J. Toneff, H. L. Johnson, S. A. Mani, P. Jones, C. C. Stephan, and J. M. Rosen, "Morphological screening of mesenchymal mammary tumor organoids to identify drugs that reverse epithelial-mesenchymal transition," *Nature Commun.*, vol. 12, no. 1, p. 4262, Jul. 2021.

[44] L. Bai, Y. Wu, G. Li, W. Zhang, H. Zhang, and J. Su, "AI-enabled organoids: Construction, analysis, and application," *Bioactive Mater.*, vol. 31, pp. 525–548, Jan. 2024.

[45] A. Petkidis, V. Andriasyan, and U. F. Greber, "Machine learning for cross-scale microscopy of viruses," *Cell Rep. Methods*, vol. 3, no. 9, Sep. 2023, Art. no. 100557.

[46] Y. Qin, J. Li, Y. Chen, Z. Wang, Y. Huang, Y. You, L. Hu, P. Hu, and F. Tan, "TransOrga: End-To-End multi-modal transformer-based organoid segmentation," in *Proc. Int. Conf. Intell. Comput.*, Jan. 2023, pp. 460–472.

[47] X. Wang, C. Wu, S. Zhang, P. Yu, L. Li, C. Guo, and R. Li, "A novel deep learning segmentation model for organoid-based drug screening," *Frontiers Pharmacol.*, vol. 13, Dec. 2022, Art. no. 1080273.

[48] E. Domènech-Moreno, A. Brandt, T. T. Lemmetyinen, L. Wartiovaara, T. P. Mäkelä, and S. Ollila, "Tellu–an object-detector algorithm for automatic classification of intestinal organoids," *Disease Models Mech.*, vol. 16, no. 3, Mar. 2023, Art. no. 049756.

[49] G. Wang, H. Ma, Y. Gao, A. Abuadbba, Z. Zhang, W. Kang, S. F. Al-Sarawi, G. Zhang, and D. Abbott, "One-to-Multiple clean-label image camouflage (OmClic) based backdoor attack on deep learning," *Knowl.-Based Syst.*, vol. 288, Mar. 2024, Art. no. 111456.

[50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[51] G. Huang, Z. Liu, G. Pleiss, L. V. D. Maaten, and K. Q. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8704–8716, Dec. 2022.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. ECCV*. Cham, Switzerland: Springer, Jan. 2016, pp. 630–645.

[54] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[55] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[59] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, Jun. 2017.

[61] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102802.

[62] G. Cirrincione, S. Cannata, G. Cicceri, F. Prinzi, T. Currieri, M. Lovino, C. Militello, E. Pasero, and S. Vitabile, "Transformer-based approach to melanoma detection," *Sensors*, vol. 23, no. 12, p. 5677, Jun. 2023.

[63] G. I. Okolo, S. Katsigiannis, and N. Ramzan, "IEViT: An enhanced vision transformer architecture for chest X-ray image classification," *Comput. Methods Programs Biomed.*, vol. 226, Nov. 2022, Art. no. 107141.

[64] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 101–121.

[65] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

GIOVANNI CICCERI received the Ph.D. degree in cyber-physical systems (CPS) from the Department of Engineering, University of Messina, Italy, in 2022. He is currently an Assistant Professor with the Department of Biomedicine, Neuroscience, and Advanced Diagnostics (BiND), University of Palermo, Italy. He collaborates on several multidisciplinary and application-oriented research projects. His research interests include the study and development of machine learning and deep learning models aimed at multi-risk and multi-objective analysis with particular reference to cyber-physical systems, embedded systems, the Internet of Things, financial engineering, vision systems, and information systems in the e-health and life sciences domains.

SEBASTIANO DI BELLA received the Ph.D. degree from the University of Catania, Italy. He was a Bioinformatician with Nerviano Medical Sciences company for almost a decade. He is currently an Associate Researcher with the Department of Precision Medicine in Medical, Surgical, and Critical Care (MePreCC), University of Palermo. He works on the development of NGS pipelines, ML/AI approach in multi-omic data, and computational analysis for biomarkers identification in oncology.

SIMONE DI FRANCO received the bachelor's degree (summa cum laude) in biotechnologies applied to feed industry from the University of Palermo, in 2007, the master's degree (summa cum laude) from the Biotechnologies for Industry and Scientific Research, in 2009, and the Ph.D. degree in immunopharmacology, focusing on p63 in breast cancer progression. In 2008, he started his research with the Laboratory of Cellular and Molecular Physiopathology under Prof. Giorgio Stassi. During his Ph.D., he spent a year with the Laboratory of Experimental Oncology and Radiobiology (AMC), Amsterdam, with Prof. Jan Paul Medema. In 2014, he completed his Ph.D. thesis on colon cancer stem cells and continued postdoctoral research with Prof. Medema and later in Italy, funded by AIRC and Fondazione Umberto Veronesi. His research focuses on the tumor microenvironment, especially the role of adipose tissue in cancer progression, with an emphasis on colorectal and breast cancer stem cells. He has studied cancer stem cell resistance to treatments and aims to explore personalized DL therapies using tumor organoids and co-cultures to test combinatorial treatments in preclinical studies.

GIORGIO STASSI began his research with the Laboratory of Immunological Medicine, University of Palermo, in 1989. In 1997, he joined the Rangos Research Center (PA), USA, under Prof. Trucco, focusing on the pathogenesis of Diabetes Mellitus. Upon returning to Italy, in 2000, he established the Laboratory of Cellular and Molecular Biology, University of Palermo with AIRC funding. In 2002, he became the Head of the Cellular and Molecular Pathophysiology Laboratory, University of Palermo. Recently, his work has been published in top journals, such as *Cell Stem Cell* and *Nature*, where he unveiled mechanisms of chemotherapy resistance in colon CSCs and proposed new therapeutic strategies. He was among the first to isolate stem cells from colon and thyroid tumors, contributing significantly to cancer research and leading to innovative therapeutic approaches. His research has focused on the role of cancer stem cells (CSCs) in epithelial tumors, such as colon, breast, and thyroid.

MATILDE TODARO was a Research Fellow with the Institute of Human Anatomy, University of Palermo, in 1988, and the Laboratory of Clinical Immunology. In 1991, she was a Visiting Assistant Professor with the Department of Immunology, Royal Liverpool Hospital, U.K., under Prof. P. M. Johnson's supervision. From 1992 to 1997, she was a Resident in endocrinology and metabolic disease with the University of Palermo. With extensive experience, she has pioneered the isolation and characterization of Cancer Stem Cells from solid tumors, especially breast cancer, shedding light on their molecular and biological features. Her success is due to key national and international collaborations, along with an advanced research team and innovative technologies. This work has contributed significantly to understanding breast cancer stem cell biology and the mechanisms behind tumor cell survival and resistance to therapy.

SALVATORE VITABILE is currently a Full Professor with the Department of Biomedicine, Neuroscience, and Advanced Diagnostics, University of Palermo, Italy. He is the co-author of more than 200 scientific papers in referred journals and conferences. He has chaired, organized, and served as a member of the organizing committee of several international conferences and workshops. His research interests include medical data processing and analysis, clinical decision support systems, specialized architecture design and prototyping, and machine and deep learning applications. He is an Associate Editor of *Human-Centric Computing and Information Sciences* journal and an Editorial Board Member of *Electronics*.

● ● ●