Discover Artificial Intelligence

Research

Optimizing cervical cancer classification using transfer learning with deep gaussian processes and support vector machines

Emmanuel Ahishakiye[1] · **Fredrick Kanobe**[2]

Received: 24 July 2024 / Accepted: 23 October 2024
Published online: 30 October 2024

Abstract

Background Cervical cancer is the fourth most frequent cancer in women worldwide. Even though cervical cancer deaths have decreased significantly in Western countries, low and middle-income countries account for nearly 90% of cervical cancer deaths. While Western countries are leveraging the powers of artificial intelligence (AI) in the health sector, most countries in sub-Saharan Africa are still lagging. In Uganda, cytologists manually analyze Pap smear images for the detection of cervical cancer, a process that is highly subjective, slow, and tedious. Machine learning (ML) algorithms have been used in the automated classification of cervical cancer. However, most of the MLs have overfitting limitations which limits their deployment, especially in the health sector where accurate predictions are needed.

Methods In this study, we propose two kernel-based algorithms for automated detection of cervical cancer. These algorithms are (1) an optimized support vector machine (SVM), and (2) a deep Gaussian Process (DGP) model. The SVM model proposed uses an optimized radial basis kernel while the DGP model uses a hybrid kernel of periodic and local periodic kernel.

Results Experimental results revealed accuracy of 100% and 99.48% for an optimized SVM model and DGP model respectively. Results on precision, recall, and F1 score were also reported.

Conclusions The proposed models performed well on cervical cancer detection and classification, and therefore suitable for deployment. We plan to deploy our proposed models in a mobile application-based tool. The limitation of the study was the lack of access to high-performance computational resources.

Keywords Kernel methods · Gaussian process models · Support vector machines · Transfer learning · Cancer prediction

1 Background

Cervical cancer ranks as the fourth most common cancer in women worldwide [1]. In 2022, more than 660,000 new cases were diagnosed. Also, in the same year, an estimated 94% of 350,000 deaths due to cervical cancer occurred in low and middle-income countries. The highest incidence and mortality rates from cervical cancer are reported in sub-Saharan Africa, Southeast Asia, and Central America. Since cervical cancer mainly affects young women, it is estimated that around 20% of children have been orphaned due to this disease [1]. Even though deaths from cervical cancer have significantly declined in Western countries, about 90% of all cervical cancer deaths worldwide occur in low- and middle-income countries. Again, in developing countries, the mortality rate from cervical cancer

✉ Emmanuel Ahishakiye, ahishema@gmail.com; Fredrick Kanobe, fkanobe@kyu.ac.ug | [1]Department of Networks, Data Science and Artificial Intelligence, Kyambogo University, Kampala, Uganda. [2]Department of Computer Science, Kyambogo University, Kampala, Uganda.

is 18 times more than in developed countries, while it accounts for 12% of female cancers worldwide [2, 3]. In Kenya, cervical cancer has a recurrence rate of 20.5 per 100,000 women and forms the primary female cancer as well as the cause of cancer growth-related death for women. It is estimated that new cases and mortality will increase by 75% by 2025 if appropriate attention and early detection methods are not applied for early treatment [4]. Uganda has had the highest cases of cervical cancer in the world at 54.8 per 100,000 women, while in New Zealand it is 5.5, and 6.6 in North America. Again, approximately 6413 women in Uganda develop cervical cancer annually, while an estimated 4301 die of the disease [5]. It is possible to cure cervical cancer if detected early [6]. Cervical cancer cannot be identified without routine screening, which involves looking for abnormal cells of the cervix that may become cancerous. Cervical cancer screening includes cervical cytology, otherwise known as a Pap test or a Pap smear; the Human Papillomavirus test; and co-testing, where the Pap smear is combined with the HPV test.

Early detection of cervical cancer through screening methods such as Pap smear tests has significantly reduced mortality rates in especially higher-income countries [7, 8]. However, manual analysis of Pap smear images remains labor-intensive, subjective, and prone to human error, particularly in regions with limited access to trained cytologists [9, 10]. This has led to a growing interest in applying machine learning (ML) and deep learning (DL) techniques to automate the classification of cervical cancer from Pap smear images. Recently, machine learning in medical imaging has gained much attention [9]. Given the capabilities for automatic feature extraction from images hierarchically, CNNs have dominated [11]. This makes them very effective in certain tasks like tumor detection and the classification of cancers [12–14]. The CNN-based architectures have been prominent in the detection of abnormalities within Pap smear and colposcopy images toward cervical cancer diagnosis [15]. On the other hand, it also belongs to the most important challenge for CNNs: overfitting a small dataset general limitation in medical imaging with respect to the scarcity of labeled data [16].

To address this issue, transfer learning has been widely adopted [17, 18]. Transfer learning leverages pre-trained models on large datasets and fine-tunes them on smaller domain-specific datasets [19]. Studies applying transfer learning to cervical cancer classification have shown improved performance, particularly when coupled with CNN architectures [20, 21]. Despite these advancements, CNNs are inherently deterministic and lack the ability to quantify uncertainty, which is critical for applications in medical diagnostics, where uncertain or borderline cases should be flagged for further investigation [22]. Beyond CNNs, other approaches such as Support Vector Machines (SVM) and Random Forests have been used in medical image classification, often with feature engineering techniques like texture analysis or histogram of oriented gradients (HOG) [23, 24]. While these models are effective on small datasets and provide interpretable results, their performance is typically inferior to that of deep learning methods in complex image classification tasks. Additionally, they lack the capability to model uncertainty, which is crucial for high-risk medical decisions.

Recent developments in machine learning methods for medical imaging include the usage of probabilistic models such as Deep Gaussian Processes [25]. Similarly, DGP extends GP to multiple layers, hence enabling complex data modeling using hierarchical representations while keeping a Bayesian framework that naturally includes uncertainty [26, 27]. The diverse applications of DGPs range from natural language processing to image classification. The probably biggest advantage of using DGPs over CNNs within a medical setting is that they are able to provide uncertainty estimates for each prediction [22]. However, the major disadvantage of using DGPs at the moment is that these models are computationally expensive and hard to train on large datasets, thus limiting their applicability unless combined with more efficient models or techniques such as transfer learning [28].

This study leverages such advances through the integration of transfer learning with two independent machine learning models, namely Deep Gaussian Processes and Support Vector Machines, to optimize cervical cancer classification from Pap smear images. Such a limitation in a small dataset is overcome by transfer learning, allowing DGP and SVM to effectively perform better when only a few labeled Pap smear images are available. The DGP model supports uncertainty estimation, which is essential in medical applications since cases with low confidence should flag further review. On the other side, the classical robust model is SVM, serving as a lower bound for comparing against other more complicated probabilistic models such as DGP in this application. Although several CNN-based methods have been proposed earlier for cervical cancer classification, this work offers unique value addition in probabilistic model evaluation and that of classical methods, such as SVM, hence filling a gap in the literature. This study has been developed for SDG 3: "*Ensure healthy lives and promote well-being for all at all ages*," within the theme of artificial intelligence for development.

1.1 The novelty and organization of the paper

The novelty of the study lies in the following aspects: (1) Transfer Learning for Enhanced Performance: Applying transfer learning in conjunction with DGP and SVM for cervical cancer classification introduces a novel approach. Transfer learning enables models to leverage pre-trained neural networks, particularly useful for medical image data, which is often limited in quantity. This transfer learning strategy helps boost the performance of both models, especially in complex diagnostic tasks. (2) Independent Evaluation of DGP and SVM: The study introduces the application of Deep Gaussian Processes (DGP) and Support Vector Machines (SVM), evaluated independently, for cervical cancer classification from Pap smear images. While both techniques are known in machine learning, using them in this specific context with transfer learning offers a fresh perspective, particularly in comparing how each performs on medical image data.

The rest of the paper is structured as follows: Sect. 2 provides an overview of Deep Gaussian Processes and Support Vector Machines. This includes subsections like 2.1 on Gaussian Processes, 2.2 on kernel functions, and 2.3 on Support Vector Machines. Section 3: Materials and Methods; Subsections 3.1: Dataset and Preprocessing techniques; 3.2: Proposed Kernel Function for the Deep Gaussian Process model; 3.3: Optimization of the SVM through the kernel function; 3.4: Proposed Models; and 3.5: Model Evaluation. The results of this study are presented in Sect. 4, with subsections on the parameters of pre-trained models, the performance of the pre-trained model on the Pap smear dataset, training versus validation accuracy, and training versus validation loss. Section 5 discusses the findings and Sect. 6 concludes this paper.

2 Literature review

2.1 Gaussian processes

The Gaussian process (GP) is a very powerful, yet structured non-parametric method applied to many statistical analysis and machine learning problems. It extends the Gaussian probability distribution into a Gaussian process [29]. GP models have been applied in Classification tasks [28]. This research covers GP classification. Essentially, a GP sets up a prior over functions and updates it to obtain a posterior over functions given the data. Although that looks like it would be hard to describe distribution over a function, actually turns out that all we need is to be able to define a distribution over the function values at a finite but arbitrary collection of points, say,$x_1, \ldots, x_N$. A Gaussian Process assumes that $p(f(x_1), ..., f(x_N))$ is jointly Gaussian with mean $\mu(x)$ and covariance $\Sigma(x)$ given by$\Sigma_{ij} = k(x_i, x_j)$, where k is a positive definite kernel function. Also, if x_i and x_j are similar, the output of the function at those points is similar [30]. The prior classification function of GP is denoted by:

$$f(x) \sim GP\big(m(x), k\big(x, x'\big)\big) \tag{1}$$

where $m(x) = \mathbb{E}[f(x)]$ represents the mean function, and $k(x, x\prime) = \mathbb{E}\big[(f(x) - m(x))(f(x\prime) - m(x\prime))\big]$ denotes the kernel or covariance function, which is positive definite [28]. A kernel function is defined as a real-valued function with two arguments, $k(x, x\prime) \in R$, for $x, x\prime \epsilon \chi$. Normally, it is symmetric, that is, $k(x, x\prime) = k(x\prime, x)$, and nonnegative, that is, $k(x, x\prime) \geq 0$ so that it can be considered as a measure of similarity [31]. More so, a function $k(x, x\prime)$ is said to be a kernel function if it is defined solely in terms of inner products in the input space [32].

Modeling with GPs depends to a great extent on the choice of kernel. Commonly used stationary kernels result in low-quality models. Deep Gaussian processes can model networks of GP nodes, and performance can be improved, which are multi-layer extensions of Gaussian processes. While DGPs are more flexible and powerful than shallow GPs, they can produce degenerate models in case the individual GP layers are not invertible, hence limiting their potential [33]. GPs can be categorized into two major types: single-layer and deep Gaussian processes. Single-layer GPs model data with a single function defined by a mean and covariance function that captures simple relationships in the data [34]. Deep Gaussian processes (DGPs), on the other hand, stack multiple layers of GPs, allowing them to model complex, hierarchical structures [35, 36]. This multi-layer approach enhances flexibility and representation power. While single-layer GPs are easier to interpret and compute, DGPs offer greater expressiveness at the cost of increased computational complexity and potential issues with non-invertible layers [37].

2.2 Kernel functions used in Gaussian processes

Several Kernel functions are used when modeling with Gaussian processes. The common Kernel functions include the periodic kernel, locally periodic kernel, squared exponential (SE) kernel, Matern kernel, and Rational Quadratic Kernel. For details, refer to [28, 31, 38, 39]. In this study, we used periodic kernels and locally periodic kernels to develop a hybrid kernel that was used in this study. This study was motivated to use the periodic and locally periodic kernels since their exclusive representation capability captures repeating patterns and local variation in the data, a very important property in medical imaging, in particular in the case of Pap smear images for cervical cancer classification [28]. In this context, periodic kernels are effective in modeling periodic or repeating structures in data. In Pap smear images, cellular structures are likely to be regularly distributed in space-for example, patterns in the cytoplasm or nucleus distribution. These types of repeating features are among the most informative priors for distinguishing cells as normal or abnormal. Locally Periodic Kernels This model integrates the strengths of periodic kernels for periodic or repeat patterns and squared exponential kernels to manage smooth nonperiodic variations [28]. The combination shall be especially effective for medical images, where periodic patterns may exist on a global scale but local variations are of prime importance for classification, including very subtle variations in cell shape, size, or texture. In Pap smear images, while there could be some structures that may present periodicity, localized variations have proved critical indicators of abnormalities, such for instance, those in cell shape. This will make a locally periodic kernel more flexible and effective in modeling both the regular structures and the local deviations in complex image data. We discuss them in detail below.

2.2.1 The periodic Kernel

The periodic kernel [28, 31, 38, 39], as defined in Eq. 2, is useful for modeling functions that exhibit exact repetition. Its parameters are straightforward:

$$\mathrm{k_{Per}}\left(\mathrm{x}, \mathrm{x}'\right) = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|\mathrm{x}-x'|/\mathrm{p}}{\ell^2}\right) \tag{2}$$

where the period p determines the interval between repetitions of the function, similar to the SE kernel, the lengthscale ℓ dictates the function's lengthscale.

2.2.2 The locally periodic Kernel

The Locally Periodic Kernel The overwhelming majority of periodic functions do not repeat exactly. We can combine a local kernel, such as the squared exponential, with our periodic kernel to add flexibility to our model [28, 31, 38, 39]. This would allow us to model only locally periodic functions where there is variation over time in the shape of the repeating component. Equation 3 gives the locally periodic kernel:

$$\mathrm{k_{LPer}}\left(\mathrm{x}, \mathrm{x}'\right) = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|\mathrm{x}-\mathrm{x}'|/\mathrm{p}}{\ell^2}\right) \exp\left(-\frac{\left(\mathrm{x}, \mathrm{x}'\right)^2}{2\ell^2}\right) \tag{3}$$

2.3 Support vector machines

Support vector machines (SVM), continue to be among the most widely used and accurate classifiers [40]. It is now a widely used method for classification tasks. SVMs are thought to be the most well-known example of kernel approaches, which use linear estimating techniques on a high dimensional feature space to handle challenging machine learning problems [40]. In this study, we considered SVM for multi-class diagnosis of cervical cancer.

Given a set of training patterns $\mathcal{T} = \left\{x_i \in \mathbb{R}^n, i = 1, \ldots, N\right\}$, and their corresponding labels from two classes $y_i \in \{-1,1\}, i = 1, \ldots, N$. The classification problem is then formulated as $y(x) = w^T\phi(x) + b$, where ϕ is the feature space

transformation and b is the linear classification bias. SVM looks for the best hyper-plane with the biggest possible margin between the closest positive and negative samples. The expression for this search is:

$$\arg\min_{w,b} \frac{1}{2} \parallel w \parallel^2, \quad subjectto : y_i\left(w^T \phi(x) + b\right) \geq 1 \tag{4}$$

In this study, we adopt the formulation by [40] that introduced a Lagrange multiplier $\alpha = \{\alpha_i\}, i = 1, \ldots, N$ that converts the formulation in Eq. (4) into a maximization problem with respect to α. Even with well-known kernels capable of representing non-linear decision boundaries, a problem with high levels of noise may become extremely challenging to solve computationally. See Sect. 3.3 for details on the optimization of the SVM.

2.4 Review of machine learning in cancer detection and classification

The study [41] conducted a systematic review of real-world applications of machine learning, such as the classification of cancer. It also covered the use of supervised, unsupervised, and reinforcement learning algorithms on medical data to classify various cancer types and predict their prognoses. It was discovered that machine learning has a great deal of promise to advance cancer diagnosis and treatment by accurately classifying different cancer types, predicting patient outcomes, and identifying prospective therapeutic targets.

To assess the risk factors of malignant cervical development, the study [42] proposed a method called CervDetect, which makes use of machine learning algorithms. CervDetect selects important features by using the random forest (RF) feature selection technique. Lastly, CervDetect employs a hybrid strategy to identify cervical cancer by fusing shallow neural networks with radiofrequency technology. With a 93.6% accuracy rate, a mean squared error (MSE) error of 0.07111, a false-positive rate (FPR) of 6.4%, and a false-negative rate (FNR) of 100%, the results demonstrate how well CervDetect diagnoses cervical cancer.

The study [13] proposed a model named Lightweight Ensemble for Brain Cancer-Grading and Classification, results revealed that the proposed model achieved 93.0% accuracy, 0.94 precision, 0.93 recall, 0.94 F1 score, and an area under the Receiver Operating Characteristic Curve (AUC-ROC) value of 0.984.

The study [43] also used federated learning and leveraging the U-Net-based model architecture in Enhancing Brain Tumor Segmentation Accuracy, the experimental results showcase the remarkable effectiveness of federated learning, significantly improving specificity to 0.96 and the dice coefficient to 0.89 with the increase in clients from 50 to 100.

The study [44] reviewed machine learning algorithms for predicting the survival of cervical cancer patients. The most popular machine learning models, according to the results, were deep learning, ensemble and hybrid learning, logistic regression, random forest, and support vector machines. Furthermore, the prediction of cervical cancer survival can be greatly influenced by the use of machine learning approaches with heterogeneous multidimensional data. It was also discovered that, despite machine learning's advantages, one of the main obstacles is still interpretability, explainability, and imbalanced datasets.

More so, the study [45] proposed a hybrid method that combined the handcrafted features and the features identified by CNN in different pathways to a new CNN in Brain Tumor Segmentation, the proposed model achieved an accuracy of 95%.

The study [21] proposed a novel approach that uses pre-trained deep neural network models, such as Alexnet, Resnet-101, Resnet-152, and InceptionV3, for feature extraction in order to meet the need for effective and intelligent screening. These models can successfully extract important information from the complex realm of Pap smear images because they have been meticulously calibrated through training on Pap smear images. Alongside the fine-tuning of these models, a variety of machine learning methods were incorporated. ResNet152 demonstrated remarkable performance, attaining an astounding accuracy rate of 98.08%.

The study [46] proposed a revolutionary cascaded strategy that intelligently supplies CNN with past information from handmade feature-based ML algorithms for brain tumor segmentation, The proposed GCNN architecture with two parallel CNNs, CSPathways CNN (CSPCNN) and MRI Pathways CNN (MRIPCNN), segmented BraTS brain tumors with high accuracy. The proposed model achieved a Dice score of 87% higher than the state of the art.

The study [47] found that while cervical cancer can be detected using a variety of approaches, new diagnostic technologies that enhance patient quality of life must be introduced by the clinical sector. These cutting-edge technologies use machine learning as a quantitative approach and spectroscopy as a qualitative one. The study also showed that the approaches and procedures that enable and hold potential as a novel screening instrument for the identification of

cervical cancer. Furthermore, it was discovered that, once taught, the algorithms can be repeated endlessly, unlike the years of training that are necessary for every single pathologist (not to mention how simple it is to make a spectroscope).

3 Methods

3.1 Dataset and preprocessing techniques

The dataset used in this study is composed of liquid-based cytology Pap smear images [48]. The dataset consists of a total of 963 images which are sub-divided into four sets representing the four classes: (malignant and pre-malignant) as NILM (Negative for Intraepithelial lesion or malignancy) with 613 images, LSIL (Low-grade intraepithelial lesions) with 163 images, HSIL (High-grade intraepithelial lesions) with 113 images, and SCC (Squamous Cell Carcinoma) with 74 images. Also, the proposed models were fine-tuned using a dataset obtained in Uganda. Data preprocessing was done on all the Pap smear images to maintain uniform quality and relevance for model training. For example, to improve the diversity of the dataset, data augmentation approaches can be employed, including rotation, flipping, and scaling. This increases the generalization capability of the models. The pixel values are finally normalized into a standard range; for example, from [0, 1], to ensure that the nature of the input provided to machine learning algorithms is always uniform. These preprocessing steps attempt to improve the quality of images and make the dataset suitable for the accurate classification of cervical cancer.

3.2 The hybrid Kernel function for the deep Gaussian process model

In this study, we proposed to use a hybrid of a periodic kernel and a locally periodic kernel. The motivation for using a hybrid kernel that combines periodic and locally periodic kernels is to model both global and local structures in the data. While the periodic kernel models repeating patterns, such as cell organization in Pap smear images, the locally periodic kernel captures smooth variations that occur locally, such as subtle changes in cell shape or texture features critical in abnormality detection. This combination enhances the model's flexibility and accuracy in classifying cervical cancer because the possible large-scale and fine-grained patterns may be learned effectively. In this section, we describe how a hybrid kernel function for use with the Deep Gaussian Process model was developed. There have been many approaches developed for creating new kernels from existing kernels [28, 31, 38, 39]. The most intuitive approach to creating new kernels from existing ones is to multiply them together, especially when they are defined over different inputs to the function. Multiplying two kernels together has some similarities with performing an AND operation. Therefore, given two valid kernels $\mathrm{K}_1\left(\mathrm{x}, \mathrm{x}'\right)$ and $\mathrm{K}_2\left(\mathrm{x}, \mathrm{x}'\right)$, their product is also a kernel, as shown in Eq. 8.

$$\mathrm{k}_1 \times \mathrm{k}_2 = \mathrm{k}_1\left(\mathrm{x}, \mathrm{x}'\right) \times \mathrm{k}_2\left(\mathrm{x}, \mathrm{x}'\right) \tag{8}$$

The proposed hybrid function in Eq. 9 is the product of a periodic kernel and a locally periodic kernel:

$$\mathrm{k}_{\mathrm{Hybrid}}\left(\mathrm{x}, \mathrm{x}'\right) = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|\mathrm{x}-\mathrm{x}'|/\mathrm{p})}{\ell^2}\right)\left[\sigma^2 \exp\left(-\frac{2\sin^2(\pi|\mathrm{x}-\mathrm{x}'|/\mathrm{p})}{\ell^2}\right)\exp\left(-\frac{\left(\mathrm{x}-\mathrm{x}'\right)^2}{2\ell^2}\right)\right] \tag{9}$$

Here, ℓ is the length scale, σ^2 is the output variance, and the weighting of large-scale and small-scale fluctuations is determined by the parameter α. Default values for these parameters were used in the experiments. For more details on kernel design, refer to the sources [28, 30, 31].

3.3 Optimization of the SVM through the Kernel function

In this section, we discuss how a radial basis kernel function was optimized. In this study, we adopt the formulation by [40] that introduced a Lagrange multiplier $\alpha = \{\alpha_i\}, i = 1, \ldots, N$ that converts the formulation in Eq. (4) into a maximization problem with respect to α. Even with well-known kernels capable of representing non-linear decision boundaries, a problem with high levels of noise may become extremely challenging to solve computationally. The ideal strategy for handling challenging issues is to introduce control parameters that permit margin limitations to be violated. This is known as the "soft-margin problem," and the dual formulation that follows can be used to define the optimization problem:

$$\max_{\alpha} D_{\gamma}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{J=1}^{N} \alpha_i \alpha_j y_i y_j k_{\gamma}(x_i, x_j) subjectto : \begin{cases} 0 \le \alpha_i \le C \forall i \\ \sum_{i=1} y_i \alpha_i = 0 \quad \forall i \end{cases} \tag{5}$$

where k_{γ} denotes the radial basis function kernel with the formulation:

$$k_{\gamma}(x, y) = \exp\left(-\gamma \parallel x - y \parallel^2\right) \tag{6}$$

The regularization term, parameter C, manages the permitted misclassification level for the training samples. Small values of C cause the optimizer to search for a hyperplane with a wide margin separation, which could lead to some points being misclassified. On the other hand, large values of C search for a smaller margin to improve the classification of all training points. Also, it was shown that for any fixed kernel k_{γ} the quantity $\max_{\alpha} D_{\gamma}(\alpha)$ is an upper bound on the miscalculation probability. The computed multipliers α_i^* and b^* allow the determination of the class of any test sample $x \in \mathbb{R}^n$ by using the formulation below:

$$f(x) = sign\left(\sum_{i \in S} y_i \alpha_i^* k_{\gamma}(x, x_i) + b^*\right) \tag{7}$$

where S is the support vectors' set of indices.

The radial basis kernel function was optimized using the kernel hyperparameter γ. The optimal value of γ is found by minimizing $\max_{\alpha} D_{\gamma}(\alpha)$ with respect to γ. Also, $D_{\gamma}(\alpha)$ is maximized with respect to α but also minimized with respect to γ. This results in a double optimization problem with the formulation below:

$$\min_{\gamma} \left(\max_{\alpha} D_{\gamma}(\alpha)\right) \quad subjectto : \begin{cases} 0 \le \alpha_i \le C \forall i = 1, \ldots, N \\ \sum_{i=1} y_i \alpha_i = 0 \quad \forall i = 1, \ldots, N \end{cases} \tag{10}$$

In this study, we adopted the minimization with respect to γ using the gradient descent approach that was earlier proposed by the study [40]. In this approach, the quantity $D_{\gamma}(\alpha)$ for the current value of α is assumed to be the suitable approximation of $\max_{\alpha} D_{\gamma}(\alpha)$. As such, the traditional steps of maximization about α are interleaved by steps of minimization concerning γ. The need for interleaving arises from the fact that attempting to determine the optimal γ value that minimizes $D_{\gamma}(\alpha)$ at every step would result in non-convergence when the recalculated value causes jumps in the reduction process. As a result, the gradient of D_{γ} with respect to D_{γ} must be computed for the gradient descent method's update step as shown below:

$$\begin{aligned} \frac{\partial D_{\gamma}(\alpha)}{\partial \gamma} &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \parallel x_i - x_j \parallel^2 \exp\left(-\gamma \parallel x_i - x_j \parallel^2\right) \\ &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \parallel x_i - x_j \parallel^2 k_{\gamma}(x_i, x_j) \end{aligned} \tag{11}$$

The gradient descent adaptation rule's update step is therefore expressed as:

$$\gamma_{t+1} = \gamma_t - \eta \frac{\partial D_{\gamma t}(\alpha)}{\partial_{\gamma t}} \tag{12}$$

where t is the time step and η is the learning rate. Also, the regularization term C is an input hyperparameter that is not optimized using the formulations above.

The gradient descend method's learning rate parameter, η, is adapted as follows: (1) The learning rate is increased by a factor $\zeta^+ : \eta = \zeta^+ \eta$ if the minimum was not reached, and (2) decreased by a factor $\zeta^- : \eta = \zeta^- \eta$ if the minimum was obtained but the algorithm did not converge, or if the actual η is too high. In this study, the validation set was used to choose the update factors, which were set at $\zeta^- = 0.1$ and $\zeta^+ = 1.01$. Similarly, the learning rate's initial value was set to $\eta = 0.01$. For details, refer to [40].

3.4 The proposed models

In this study, we were mostly interested in the application and fine-tuning of two popular machine learning models i.e. the Deep Gaussian Processes (DGP) and Support Vector Machines (SVM) rather than developing new models from scratch. We also developed a hybrid kernel function using periodic and locally periodic kernels for DGP discussed in Sect. 3.2, and we optimized the RBF kernel for SVM discussed in Sect. 3.3, to improve the model performance for catching more complex patterns in cervical cancer classification. Moreover, through the process of transfer learning, we integrated a pre-trained InceptionV3 backbone to make use of the feature extraction capabilities of the network. In this way, we fine-tuned the existing InceptionV3 architectures to perform optimally on our dataset. Although no new model architectures were developed in this work, the approach presented in this study is an adaptation and optimization of already existing techniques for the task of cervical cancer detection using Pap smear images. All implementations were carried out using Google Colab. The Deep Gaussian Process (DGP) model was implemented with Python libraries GPflow [49] and GPflux [50], while the Support Vector Machine (SVM) model was developed using the Keras library.

3.5 Hyperparameter optimization of the proposed models

The hyperparameters of the SVM were optimized as follows. In this study, an optimized radial basis function (RBF) kernel was used for the SVM model. The hyperparameters of the RBF which are the kernel coefficient (gamma) and regularization parameter (C) were optimized using a grid search approach. a grid search over *C* of [0.1, 1, 10, 100] was done and a value of 10 was found to be optimal during the experimental results. Also, fivefold cross-validation was done to ensure generalization across training data. For the Gamma (Kernel Coefficient), A grid search over γ values of [0.01,0.1,1] was done. The value $\gamma = 0.1$ was found to be optimal. Therefore, the regularization parameter $C = 10$ and the Gamma value $\gamma = 0.1$ were used in all our experiments.

The DGP model in the study uses a hybrid kernel that combines periodic and local periodic kernels to handle the non-stationarity of the data. The Bayesian optimization optimization was used to determine the optimal values for the hyperparameters. The Length Scale of 2 for the proposed kernels was used, the noise variance was 0.01, and the number of layers of the DGP model was 2. Also, stochastic gradient descent was used for optimization and the Adam optimizer with a learning rate of 0.001 was used.

3.6 Model evaluation

The proposed models were assessed based on accuracy, cross-entropy loss, precision, recall, F1-score, and the loss function. Accuracy is measured by the ratio of correctly predicted outcomes to the total number of predictions [51]. Since accuracy provides a quick measure of a model's performance and is particularly effective for classification problems, it was selected for evaluating this task. Also, the cross-entropy metric measures the model's error and dissimilarity between predicted and actual values. The performance metrics are calculated using the formulas below:

$$ccuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall, Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

Fig. 1 The pseudo-code of the proposed model

1: Import necessary libraries
Import GPflow, TensorFlow, Keras, sklearn, numpy, pandas, matplotlib
2: Load and preprocess the dataset
Load cervical cancer dataset
3: Preprocess the dataset
Resize images to a standard size, Normalize image pixel values, Perform data augmentation
4: Define the Transfer Learning Model
Load a pre-trained model (InceptionV3) without the top layers
5: Fine-tune the Transfer Learning Model
Train the model on train_data
Validate the model on val_data
6: Extract Features using the Trained Model
Remove the final classification layers from the model
Pass the data through the model to get feature representations
6: Train the Kernel Methods (SVM, Deep Gaussian Processes)
Train a Support Vector Machine (SVM) with the extracted features
Train a Deep Gaussian Process model with the extracted features
7: Evaluate the Models
Make predictions on the test_data
Calculate evaluation metrics (accuracy, precision, recall, F1 score, AUC)
Generate a confusion matrix

where *TP*, *TN*, *FP*, *andFN* stand for true positive, false positive, and negative, respectively. In this study, the implementations were performed using Python 3.12.4, TensorFlow 2.16.0, and Keras 3.4.1. The work was conducted on a Windows 11 laptop equipped with an Intel Core i7 processor running at 2.50 GHz, 16 GB of RAM, and a 2 GB NVIDIA GeForce MX150 GPU.

3.7 The pseudo code of the proposed models

The pseudocode in Fig. 1 describes a workflow that applies transfer learning and kernel-based methods, notably SVM and Deep Gaussian Processes, to classify cervical cancer from Pap smear images. Firstly, the required libraries are imported followed by the preprocessing. A pre-trained InceptionV3 model loaded and fine-tuned on the cervical cancer dataset, adapting its learned features for this particular task. In this fine-tuned model, the final classification layers of the InceptionV3 are removed after which data is passed through the model to extract feature representations. Both Support Vector Machine and Deep Gaussian Processes models are trained on extracted features and further evaluated on test data based on accuracy, precision, recall, F1 score, AUC, and confusion matrices.

4 Experimental results

4.1 Pre-trained transfer learning model parameters

In this study, we compare the performance of the two proposed models, these are (1) a model that integrates a pretrained transfer learning model (InceptionV3) with SVM, and (2) a model that integrates a pretrained transfer learning model (InceptionV3) with DGP. Both models were built by first leveraging a pretrained Deep learning model to extract high-level features from the Pap smear images. This transfer learning model was then fine-tuned on the cervical cancer dataset to adjust to domain-specific features. After feature extraction, the deep network's final layer was replaced with a Support Vector Machine (SVM), and DGP classifiers respectively for (1) and (2) models above. This required the determination of model parameters, including total parameters, trainable parameters, and non-trainable parameters. As shown in Table 1, the most total and trainable parameters were in ResNet50 among all these pre-trained models, and that of MobileNetV2

Table 1 Transfer learning model parameters

Model	Trainable parameters	Non-trainable parameters	Total parameters
InceptionV3	21,802,784	21,768,352	34,432
MobileNetV2	2,257,984	2,223,872	34,112
ResNet50	23,534,592	53,120	23,587,712
VGG16	14,714,688	14,714,688	0
DenseNet201	18,092,928	229,056	18,321,984
VGG19	20,024,384	20,024,384	0

Table 2 Performance of the pre-trained models

Model	Training accuracy (%)	Validation accuracy (%)	Training loss	Validation loss
VGG16	83.45	72.75	0.82	0.90
VGG19	82.66	71.14	0.63	0.94
ResNet50	70.50	64.52	0.88	2.20
InceptionV3	98.80	85.26	0.22	2.58
DenseNet201	95.72	82.33	0.35	2.73
MobileNetV3	97.81	88.84	0.12	2.95

was the lowest. Both VGG16 and VGG19 cleared with no non-trainable parameters, whereas DenseNet201 came out to be the model with the least non-trainable parameters.

4.2 Performance of the pre-trained transfer learning models

Some of the pre-trained transfer learning models fine-tuned on the dataset utilized in this study for cervical cancer are shown in Table 2. These models have been pre-trained and hence these can leverage feature representations learned on such a large and diverse dataset. In order for these learned features of the model to adapt to this specific classification task, the final layers of each model were replaced and further retrained using the cervical cancer dataset. Fine-tuning entails updating, especially the higher layers, of model parameters while keeping fixed or with small learning rates those earlier layers that pick out more general features such as edges and textures. The authors measured the training accuracy, validation accuracy, training loss, and validation loss for each of the models after subjecting them to fivefold cross-validation to assess their generalization on unseen data. Fine-tuning allowed the models to specialize in feature detection relevant to the diagnosis of cervical cancer using Pap smear images. InceptionV3 showed an overall best performance and was therefore selected as the model of choice in this study.

4.3 Experimental results of the proposed models

In this section, we report results obtained from our two proposed models. Our two proposed models are (1) the DGP model with a hybrid kernel function presented in Sect. 3.2; and (2) an SVM with an optimized kernel function presented in Sect. 3.3. The proposed models were integrated with an InceptionV3 pre-trained transfer learning model where each of the proposed models works as the prediction layer. In this work, the authors performed a fine-tuning of an InceptionV3 model that was pre-trained on ImageNet, on the dataset of Pap smears to perform cervical cancer classification. First, they removed the final fully connected layers of the InceptionV3 model, which are specific to ImageNet classes. Then, they kept previous layers, since these layers capture general features of the visual appearance, like edges and textures, which could be easily transferred to medical imaging tasks. This model was then fine-tuned-retraining the latter layers of InceptionV3 on the Pap smear images. The latter layers were initialized with a small learning rate as a way of preventing overwriting of pre-trained weights and allowing these layers to adapt to unique features of the Pap smear dataset. The network was then trained on selected classes of cervical cancer, which were normal, pre-cancerous, and cancerous classes. The network applied augmentation techniques to develop robustness in the model through rotation, flipping, and scaling. Finally, the fully connected layers were replaced by

Table 3 Classification report for the DGP model

Classification report:	Precision	Recall	f1-score	Support
High squamous intra-epithelial lesion	1.00	0.97	0.98	33
Low squamous intra-epithelial lesion	1.00	1.00	1.00	22
Negative for Intraepithelial malignancy	1.00	1.00	0.97	123
Squamous cell carcinoma	0.94	1.00	0.97	15
Accuracy			0.99	193
Macro avg	0.98	0.99	0.99	193
Weighted avg	1.00	0.99	0.99	193

Table 4 Classification report for the SVM model

Classification report:	Precision	Recall	f1-score	Support
High squamous intra-epithelial lesion	1.00	1.00	1.00	33
Low squamous intra-epithelial lesion	1.00	1.00	1.00	22
Negative for Intraepithelial malignancy	1.00	1.00	1.00	123
Squamous cell carcinoma	1.00	1.00	1.00	15
Accuracy			1.00	193
Macro avg	1.00	1.00	1.00	193
Weighted avg	1.00	1.00	1.00	193

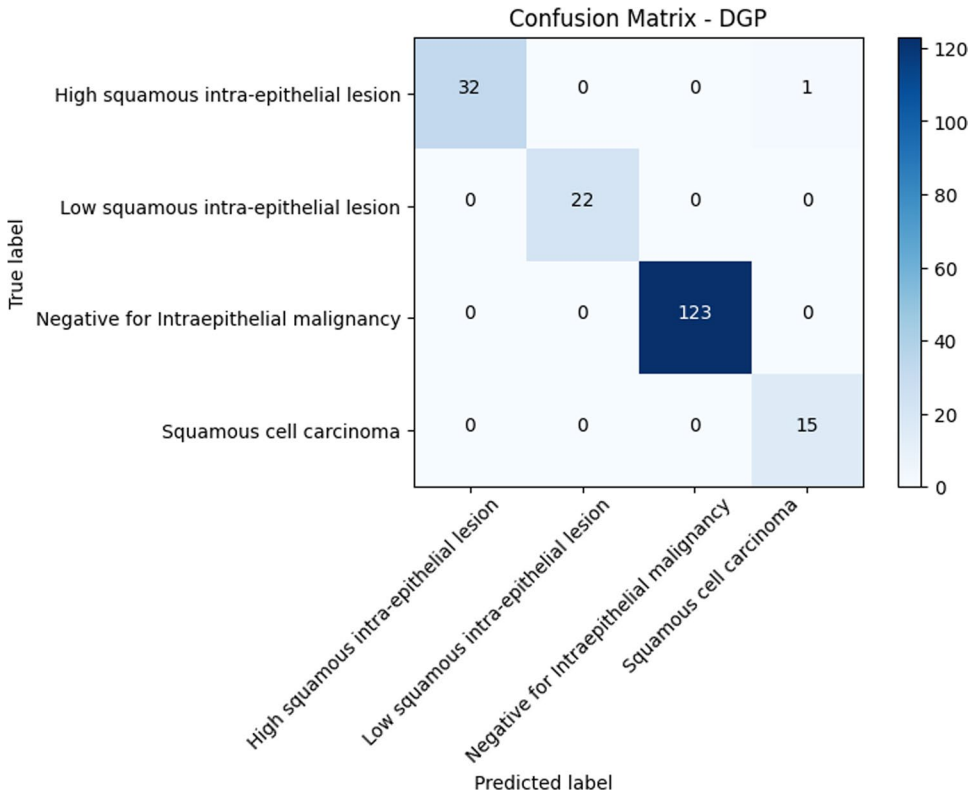

Fig. 2 The confusion Matrix for the DGP model

a classifier (SVM or DGP) suitable for the task of cervical cancer, and retraining of the whole model on the target dataset was performed to allow the pre-trained weights to adapt to the specific patterns of the Pap smear images while preserving powerful feature extraction capabilities learned from the large-scale ImageNet data. Table 3 below shows the classification report for the DGP model, and Table 4 for the optimized SVM model. The experimental results indicated that the optimized SVM model was better with respect to the classification reports given in Tables 3 and 4.

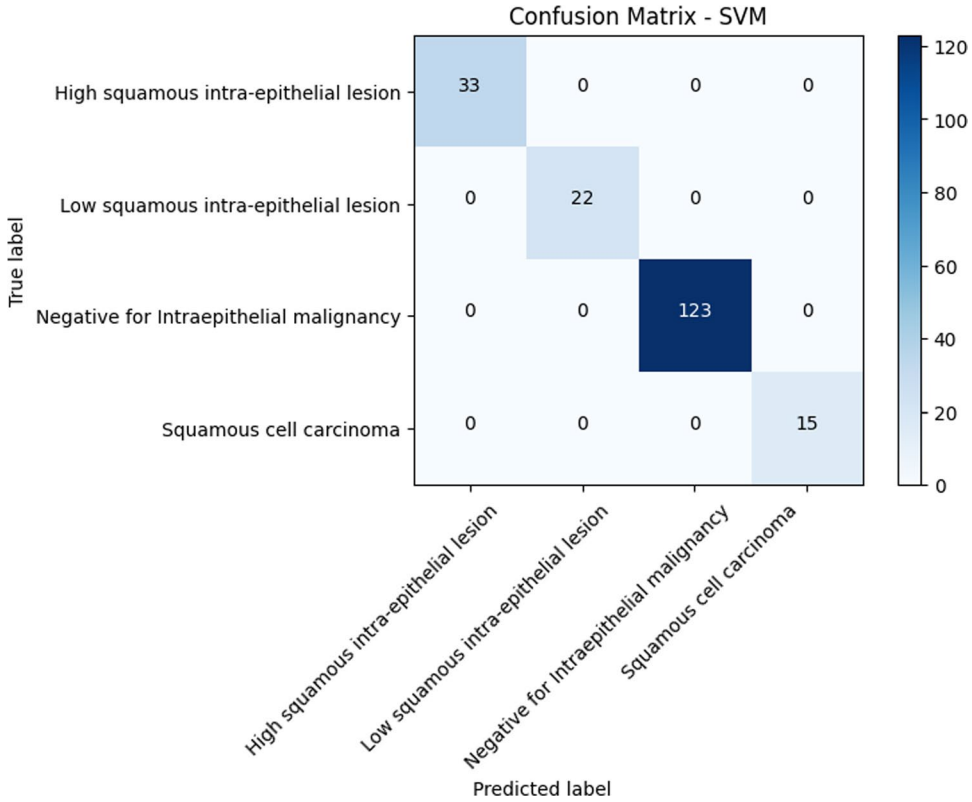

Fig. 3 The confusion Matrix for the SVM model

The following results on the confusion matrix also show the performance of the proposed models. The results show that the DGP model accurately classified all the categories except only one instance where a high squamous intra-epithelial lesion was classified as squamous cell carcinoma as shown in Fig. 2. The optimized SVM model correctly classified all the instances as shown in Fig. 3.

5 Discussion of the findings

In this study, we proposed two Kernel-based algorithms i.e. a DGP model and an optimized SVM model. The DGP model used a hybrid kernel function discussed in Sect. 3.2 while the SVM model used an optimized radial basis function discussed in Sect. 3.3. We first tested several pre-trained models of transfer learning with different performances. This study considers six models: VGG16, VGG19, ResNet50, InceptionV3, DenseNet201, and MobileNetV2. Among these, InceptionV3 performed better, and thus in this task, we adopt it as the transfer learning model with our proposed models. The transfer learning approach has been used in related studies such as early-stage brain tumor detection [52] and results reveal performance improvements. The accuracy results in this study showed that our proposed optimized SVM model, along with the pre-trained model, achieved an accuracy of 100%, whereas the DGP model, along with the pre-trained model, had an accuracy of 99.48%. More results are shown in Table 5. However, an optimized SVM model had an overall better performance when compared to the DGP model.

In comparison to the related works, the study [53] used deep convolutional neural networks and achieved an overall accuracy of 94% while the study [54] used convolutional neural networks and achieved an accuracy of 93.3%. However, most of the studies in the literature did not report their results on other performance metrics like precision, recall, and F1 score. Studies that did not use image data to detect and predict cervical cancer were not considered in this study while discussing

Table 5 The summary of the results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
DGP	99.48	99.51	99.48	99.49
SVM	100.00	100.00	100.00	100.00

our findings. To further contextualize the performance of the two models, we extended the comparison to transformer-based architectures. This model class has recently reported state-of-the-art results in various medical imaging tasks due to their intrinsic ability to learn long-range dependencies in images. A specific approach combining Low-Rank Adaptation with the Vision Transformer model, for digital pathology classification, has been proposed in [55], reaching an accuracy as high as 98.9%. The work in [56] proposed a model, ViT-PSO-SVM, to better the results of the ViT by effectively integrating it with PSO and SVM, after which the proposed model achieved 99.112% accuracy. Though these transformer-based models compete well [57], they have very high computational overhead; hence, it can limit their deployment to real clinical settings, especially in resource-constrained environments. On the other hand, DGP and SVM are more computationally efficient while still offering robust performance metrics, hence their practicality for clinical use.

5.1 Mitigating overfitting in SVM and DGP

Optimized RBF kernel handled overfitting in the SVM model. The selection of the optimal value for both the regularization parameter, C, and the kernel coefficient, γ, was determined through grid search cross-validation. This includes the search over $C = [0.1,1,10]$ and $\gamma = [0.01,0.1,1]$. It selected $C = 10$ and $\gamma = 0.1$, considered to be an optimal balance between the model complexity and accuracy without risking overfitting. In this regard, the regularization parameter, C, is used to find a good trade-off between the maximization of the margin and the correct classification of the training points. A higher value of C emphasizes fewer misclassifications. The value of γ in the RBF kernel controls the influence of a single training point; smaller values of γ will make the decision boundary smoother, hence reducing overfitting. To ensure further that our SVM model generalizes well to new data, we applied fivefold cross-validation across the Pap smear dataset. This was the approach followed: dividing the data into five subsets of training on four and testing on one, then cycling through all. This will decrease variance, decrease bias, and give the ability to generalize well. Moreover, tests on the validation set were performed separately to check whether the selected hyperparameters perform well on sets other than the training set.

To address the issue of overfitting in DGP, we utilized a hybrid kernel that was comprised of periodic and local periodic within the model to enable the model to capture the global and local periodicities in Pap smear images, which are so critical in cervical cancer classification. We further used Bayesian optimization to determine that the kernel length scale of 2 and noise variance of 0.01 provide a good trade-off between model complexity and accuracy. Bayesian regularization helped to prevent overfitting by placing priors on the model parameters, thereby not being too confident in high-dimensional data. We are quite sure the model will be robust; thus, we implemented fivefold cross-validation and set an early stopping criterion for performance checks during training.

5.2 Computational resource requirements, integration with existing medical workflows, and scaling for large datasets

To address the challenge of the demand for computational resources, we resorted to using cloud-based solutions: Google Collab and GPUs during training. Besides that, we are going to suggest deploying this model on Amazon Web Services with parallel processing. The approach we are going to propose would reduce the computational load during inference; thus, it could make an application feasible in real-time clinical use. Also, models can work much more efficiently on limited hardware, such as edge devices, which can be often utilized within medical settings.

In developing this, Integration with the Existing Medical workflow, models were designed to be interoperable-out-of-the-box according to industry standards such as HL7 FHIR and DICOM. These standards will easily enable our models to interface with the present medical imaging and record-keeping systems, thus easily deploying them in hospitals and clinics.

We rely on transfer learning for handling large-scale datasets, where large pre-trained models are used, requiring only fine-tuning on domain-specific Pap smear images. This reduces the need for large computational resources and hastens model deployment in resource-constrained environments. In addition, we further propose online learning for continuous model updates to adapt to changing data distributions and clinical environments.

6 Conclusions

The study proposed two Kernel-based algorithms. The two algorithms are an optimized SVM model and a DGP model. The SVM model used an optimized radial basis kernel function while the DGP model used a hybrid kernel. Also, the InceptionV3 transfer learning model was used together with our proposed models. Experimental results revealed that

the optimized SVM model performed better than the DGP model using the accuracy, precision, recall, and F1 score metrics. However, the optimized SVM model exhibited overfitting issues while the DGP model was robust to overfitting. The main limitation of this study was the insufficient computational resources for rapid model training. All experiments were conducted on a standard laptop, which caused some experiments to take several minutes to complete before producing results.

6.1 Future research directions

We propose applying the developed methods to other forms of cancer detection such as breast or lung cancer and extending the work to a wider range of medical imaging datasets (e.g., MRI, CT scans), which would demonstrate the broader applicability of the models and provide a foundation for improving diagnostic tools across various medical fields. Also, we propose to develop an ensemble model that combines the strengths of Deep Gaussian Processes (DGP) and Support Vector Machines (SVM) to create hybrid approaches that may further improve classification performance and robustness.

Author contributions EA and FK designed the study. EA wrote the manuscript. FK edited and extensively reviewed the manuscript. All the authors approved the manuscript.

Funding We acknowledge the financial support from Kyambogo University's competitive grants.

Data availability The data utilized in this study is sourced from the Mendeley Data "Liquid-based cytology pap smear images for multi-class diagnosis of cervical cancer," which is publicly accessible at the following URL: https://data.mendeley.com/datasets/zddtpgzv63/4. https://doi.org/10.17632/zddtpgzv63.4. Accessed 15-April-2024.

Code availability Not applicable.

Declarations

Competing interests The authors declare no competing interests.

References

1. WHO. Cervical cancer. https://www.who.int/news-room/fact-sheets/detail/cervical-cancer. Accessed 14 May 2024.
2. Cohen PA, Jhingran A, Oaknin A, Denny L. Seminar cervical cancer. Lancet. 2019;393(10167):169–82. https://doi.org/10.1016/S0140-6736(18)32470-X.
3. Manikandan S, Behera S, Naidu NM, Angamuthu V, Mohammed OFB, Debata A. Knowledge and awareness toward cervical cancer screening and prevention among the professional college female students. J Pharm Bioallied Sci. 2019. https://doi.org/10.4103/JPBS.JPBS_21_19.
4. Oketch SY, et al. Perspectives of women participating in a cervical cancer screening campaign with community-based HPV self-sampling in rural western Kenya: a qualitative study. BMC Womens Health. 2019. https://doi.org/10.1186/s12905-019-0778-2.
5. Black E, Hyslop F, Richmond R. Barriers and facilitators to uptake of cervical cancer screening among women in Uganda: a systematic review. BMC Womens Health. 2019;19(1):1–12. https://doi.org/10.1186/s12905-019-0809-z.
6. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. Biomed Eng Online. 2019;18(1):1–22. https://doi.org/10.1186/s12938-019-0634-5.
7. Casas CPR, et al. Cervical cancer screening in low- and middle-income countries: a systematic review of economic evaluation studies. Clinics. 2022. https://doi.org/10.1016/j.clinsp.2022.100080.

8. Basoya S, Anjankar A. Cervical cancer: early detection and prevention in reproductive age group. Cureus. 2022. https://doi.org/10.7759/cureus.31312.
9. Lee YM, Lee B, Cho NH, Park JH. Beyond the microscope: a technological overture for cervical cancer detection. Diagnostics. 2023;13(19):1–17. https://doi.org/10.3390/diagnostics13193079.
10. Hou X, Shen G, Zhou L, Li Y, Wang T, Ma X. Artificial intelligence in cervical cancer screening and diagnosis. Front Oncol. 2022;12:1–13. https://doi.org/10.3389/fonc.2022.851367.
11. Bakator M. Deep learning and medical diagnosis : a review of literature. Multimodal Technol Interact. 2018. https://doi.org/10.3390/mti2030047.
12. Alshuhail A, et al. Refining neural network algorithms for accurate brain tumor classification in MRI imagery. BMC Med Imaging. 2024;24(1):1–20. https://doi.org/10.1186/s12880-024-01285-6.
13. Ullah F, et al. Evolutionary model for brain cancer-grading and classification. IEEE Access. 2023;11:126182–94. https://doi.org/10.1109/ACCESS.2023.3330919.
14. Jiang X, Hu Z, Wang S, Zhang Y. Deep learning for medical image-based cancer diagnosis. Cancers (Basel). 2023. https://doi.org/10.3390/cancers15143608.
15. Alsubai S, et al. Privacy preserved cervical cancer detection using convolutional neural networks applied to pap smear images. Comput Math Methods Med. 2023;2023:1–8. https://doi.org/10.1155/2023/9676206.
16. Ahishakiye E, Van Gijzen MB, Tumwiine J, Wario R, Obungoloch J. A survey on deep learning in medical image reconstruction. Intell Med. 2021. https://doi.org/10.1016/j.imed.2021.03.003.
17. Sharma C, Parikh S. Comparison of CNN and Pre-trained models: a study. no. April, 2022. https://www.researchgate.net/publication/359850786_Comparison_of_CNN_and_Pre-trained_models_A_Study. Accessed 25 Jun 2024
18. Iman M, Rasheed K, Reza H. A review of deep transfer learning and recent advancements. Technologies. 2023;11(2):1–14. https://doi.org/10.3390/technologies11020040
19. Cui Y, Song Y, Sun C, Howard A, Belongie S. Large scale fine-grained categorization and domain-specific transfer learning. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2018. https://doi.org/10.1109/CVPR.2018.00432.
20. Abinaya K, Sivakumar B. A deep learning-based approach for cervical cancer classification using 3D CNN and vision transformer. J Imaging Inform Med. 2024;37(1):280–96. https://doi.org/10.1007/s10278-023-00911-z.
21. Mathivanan SK, Francis D, Srinivasan S, Khatavkar V, Karthikeyan P, Shah MA. Enhancing cervical cancer detection and robust classification through a fusion of deep learning models. Sci Rep. 2024;14(1):1–14. https://doi.org/10.1038/s41598-024-61063-w.
22. Dutordoir V, van der Wilk M, Artemev A, Hensman J. Bayesian Image Classification with Deep Convolutional Gaussian Processes. In Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. 2020. http://arxiv.org/abs/1902.05888.
23. Abubakar H, Al-Turjman F, Ameen ZS, Mubarak AS, Altrjman C. A hybridized feature extraction for COVID-19 multi-class classification on computed tomography images. Heliyon. 2024;10(5):e26939. https://doi.org/10.1016/j.heliyon.2024.e26939.
24. Albadr MAA, Tiun S, Ayob M, Al-Dhief FT, Omar K, Hamzah FA. Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection. PLoS ONE. 2020;15(12):1–28. https://doi.org/10.1371/journal.pone.0242899.
25. Lopez-Perez M, Morales-Alvarez P, Cooper LAD, Molina R, Katsaggelos AK. Deep Gaussian processes for classification with multiple noisy annotators. application to breast cancer tissue classification. IEEE Access. 2023;11:6922–34. https://doi.org/10.1109/ACCESS.2023.3237990.
26. Kumar V, Singh V, Srijith PK, Damianou A. Deep Gaussian processes with convolutional Kernels. 2018. http://arxiv.org/abs/1806.01655.
27. Salembeni H. Deep Gaussian processes: advances in models and inference. Imperial College London. 2020. https://spiral.imperial.ac.uk/handle/10044/1/81669.
28. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Int J Neur Syst. 2006;14(2). https://doi.org/10.1142/S0129065704001899.
29. Rasmussen CE. Gaussian processes in machine learning. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3176. Berlin: Springer; 2004. p. 63–71. https://doi.org/10.1007/978-3-540-28650-9_4
30. Bishop CM. Pattern recognition and machine learning. 1st ed. New York: Springer-Verlag; 2006.
31. Murphy KP. Machine learning: a probabilistic perspective. London: The MIT Press; 2012.
32. Wang J. An Intuitive Tutorial to Gaussian Processes Regression. Kingston, ON K7L 3N6 Canada, 2021. http://arxiv.org/abs/2009.10862.
33. Blomqvist K, Kaski S, Heinonen M. Deep convolutional Gaussian processes. In: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11907. Cham: Springer International Publishing; 2020. p. 582–97. https://doi.org/10.1007/978-3-030-46147-8_35
34. Hensman J, Fusi N, Lawrence ND. Gaussian processes for big data. Uncertain Artif Intell. 2013. https://doi.org/10.1016/S0074-7696(01)08005-6.
35. Dunlop MM, Girolami MA, Stuart AM, Teckentrup AL. How deep are deep Gaussian processes? J Mach Learn Res. 2018;19:1–46. https://doi.org/10.5555/3291125.3309616.
36. Damianou AC, Lawrence ND. Deep Gaussian Processes. J Mach Learn Res. 2013;31:207–215. http://proceedings.mlr.press/v31/damianou13a.pdf.
37. Sauer A, Cooper A, Gramacy RB. Vecchia-approximated deep Gaussian processes for computer experiments. J Comput Graph Stat. 2023;32(3):824–37. https://doi.org/10.1080/10618600.2022.2129662.
38. Murphy KP. Probabilistic machine learning: advanced topics. Cambridge: The MIT Press; 2022.
39. Duvenaud DK. Automatic Model Construction with Gaussian Processes. PhD Thesis, Univ. Cambridge, no. June. 2014; 144. https://www.repository.cam.ac.uk/handle/1810/2472810A. https://www.cs.toronto.edu/~duvenaud/thesis.pdf.
40. Thurnhofer-hemsi K, López-rubio E, Molina-cabello MA. Radial basis function kernel optimization for support vector machine classifiers. 2020. https://arxiv.org/pdf/2007.08233
41. Yaqoob A, Musheer Aziz R, Verma NK. Applications and techniques of machine learning in cancer classification: a systematic review. Human Centric Intell Syst. 2023;3(4):588–615. https://doi.org/10.1007/s44230-023-00041-3.

42. Mehmood M, Rizwan M, Gregus ml M, Abbas S. Machine learning assisted cervical cancer detection. Front Public Heal. 2021;9:1–14. https://doi.org/10.3389/fpubh.2021.788376.
43. Ullah F, Nadeem M, Abrar M, Amin F, Salam A, Khan S. Enhancing brain tumor segmentation accuracy through scalable federated learning with advanced data privacy and security measures. Mathematics. 2023. https://doi.org/10.3390/math11194189.
44. Rahimi M, Akbari A, Asadi F, Emami H. Cervical cancer survival prediction by machine learning algorithms: a systematic review. BMC Cancer. 2023;23(1):1–10. https://doi.org/10.1186/s12885-023-10808-3.
45. Ullah F, et al. Brain tumor segmentation from MRI images using handcrafted convolutional neural network. Diagnostics. 2023;13(16):1–15. https://doi.org/10.3390/diagnostics13162650.
46. Ullah F, Nadeem M, Abrar M. Revolutionizing brain tumor segmentation in MRI with dynamic fusion of handcrafted features and global pathway-based deep learning. KSII Trans Internet Inf Syst. 2024;18(1):105–25. https://doi.org/10.3837/tiis.2024.01.007.
47. Meza Ramirez CA, Greenop M, Almoshawah YA, Martin Hirsch PL, Rehman IU. Advancing cervical cancer diagnosis and screening with spectroscopy and machine learning. Expert Rev Mol Diagn. 2023;23(5):375–90. https://doi.org/10.1080/14737159.2023.2203816.
48. Hussain E. Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer. Mendeley Data. 2019. https://doi.org/10.17632/zddtpgzv63.4.
49. Matthews DG, et al. GPflow: a Gaussian process library using TensorFlow. J Mach Learn Res. 2017;18(40):1–6.
50. Dutordoir V et al. GPflux: a library for deep Gaussian processes. 2021. http://arxiv.org/abs/2104.05674
51. Li J. Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? PLoS ONE. 2017;12(8):1–16. https://doi.org/10.1371/journal.pone.0183250.
52. Md Ashafuddula NI, Islam R. ContourTL-Net: contour-based transfer learning algorithm for early-stage brain tumor detection. Int J Biomed Imaging. 2024. https://doi.org/10.1155/2024/6347920.
53. Shanthi PB, Faruqi F, Hareesha KS, Kudva R. Deep convolution neural network for malignancy detection and classification in microscopic uterine cervix cell images. Asian Pacif J Cancer Prev. 2019;20(11):3447–56. https://doi.org/10.31557/APJCP.2019.20.11.3447.
54. Wu M, Yan C, Liu H, Liu Q, Yin Y. Automatic classification of cervical cancer from cytological images by using convolutional neural network. Biosci Rep. 2018;38(6):1–9. https://doi.org/10.1042/BSR20181769.
55. Hong Z, Xiong J, Yang H, Mo YK. Lightweight low-rank adaptation vision transformer framework for cervical cancer detection and cervix type classification. Bioengineering. 2024. https://doi.org/10.3390/bioengineering11050468.
56. AlMohimeed A, Shehata M, El-Rashidy N, Mostafa S, Samy Talaat A, Saleh H. ViT-PSO-SVM: cervical cancer predication based on integrating vision transformer with particle swarm optimization and support vector machine. Bioengineering. 2024;11(7):729. https://doi.org/10.3390/bioengineering11070729.
57. Vaswani A et al. Retention Is All You Need. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017; 4752–4758. https://doi.org/10.1145/3583780.3615497.