Advancing Healthcare AI Education Through Cloud Computing: Benchmarking AWS vs. GCP

1st Ahmed Elghareeb
Electrical and Computer Engineering
Iowa State University
Ames, IA
aghaareb@iastate.edu

2nd Ibrahim Abdelmawla
Electrical and Computer Engineering
Iowa State University
Ames, IA
ibrahim4@iastate.edu

3rd Ashraf Gaffar
Electrical and Computer Engineering
Iowa State University
Ames, IA
gaffar@iastate.edu

***Abstract*—This research-to-practice full paper represents our work to comprehensively analyze cloud computing benchmarking with integrated AI-driven healthcare projects. As time progresses, the need to migrate a variety of system applications to the cloud has grown significantly. Yet, choosing a particular cloud service provider can be troublesome and proposes several challenges, whether computational or financial, depending on the domain of the application. Thus, the reliance on cloud platform benchmarking has become an essential skill engineers must develop in order to make informed and efficient decisions. Healthcare applications, in particular, constitute an excellent from-ground-to-cloud example that we needed to analyze and study, especially from the perspective of Artificial Intelligence (AI) applications. At the moment, there is high demand for AI applications in the domain of healthcare seeking higher performance while lowering the costs of the services. Nonetheless, the healthcare domain is considered a tough field to experience and experiment to new graduates and proposes challenges to successfully cloudify the designated system applications. Therefore, we developed an effectual pedagogical approach capable of addressing those stacked challenges and preparing a new generation of competent engineers. Our work is designed to implement multiple stages sequentially and individually. The first stage commences by introducing students to various concepts as a foundation for their core work, covering Machine Learning (ML) basics, healthcare domain knowledge basics, finding a solution via ground and cloud computation, and benchmarking. Students then attempted to address a chosen healthcare problem using AI techniques. Finally, they solve the proposed healthcare challenge and successfully benchmark the cloud platforms of both Google Cloud Platform (GCP) and Amazon Web Services (AWS). Students, as instructed, relied on concepts of metrics, measurements, criteria, and their disparities to accurately assess each cloud. This paper efficiently documents and analyzes the results of each proposed stage based on the represented students' deliverables and the supervisors' feedback to reflect the effectiveness of the suggested teaching approach.**

***Index Terms*—Artificial Intelligence, Machine Learning, Cloud Computing, Cloudification, Benchmarking, GCP, AWS, Healthcare, Education, Undergraduate.**

I. Introduction

The current era of modern technology is marked by rapid advancements across virtually every aspect of our lives, driven largely by new tools and technologies, including AI-powered clouds. As vital as healthcare services can be given the growing population and medical challenges, they escalate the need to revolutionize the technology used in order to maintain and improve the quality of the services and lower, or at least maintain the financial requirements. For instance, due to the COVID-19 pandemic, the need for automated and efficient systems for medical institutions rose significantly, as doctors were in deep need of artificially intelligent helping hands to make fast and accurate diagnoses, especially given the indefinitely increasing number of patients. In other words, the world is adopting the digitization of different fields with the power of fast, reliable AI computations, and the healthcare sector is a crucial environment for the same cause. At the moment, governments, companies, hospitals, and universities are directing their focus to the power of cloud computation since it holds a considerably high computational power and security measures and constitutes a convenient solution for systems deployment. The flexibility, scalability, and cost-effectiveness of cloud platforms make them indispensable for managing complex systems and processing vast amounts of data. Clouds feature streamlined operations, advanced security measures, and improved service delivery to end users, which make them a great asset for governments. Universities are fitting environments that leverage the use of cloud resources for both research purposes and the management of educational processes. Medical institutions and hospitals have turned to cloud computing to utilize its power at lower costs in patient data management, enhancing the efficiency of services, and the deployment of cutting-edge applications. However, although migrating systems from ground computation to the cloud is a pivotal fruitful process, it is deemed to be challenging. There exist many cloud service providers, and each offers different hardware and software capabilities for different price tags. Subsequently, selecting the appropriate platform requires a thorough consideration of computational and financial aspects, as much as considering the domain for which this platform is used since a domain's distinctive use would highlight performance and efficiency differences amongst different development environments. On that account, the skill of benchmarking and attentive comparisons of clouds has become crucial for both engineers and decision-makers. Benchmarking provides the needed insights into accurate and planned decisions regarding which cloud environment meets the dedicated application requirements and is better than the other. Therefore, it ensures scalability, cost efficiency, and performance optimality.

Generally, modern AI is one of the most viral and fundamental sciences specifically approached by numerous institutions, whether professionally or academically, due to its crucial role in scientific and non-scientific communities and non-stopping fast-growing innovations. Therefore, having a solid pedagogical paradigm and approach to eradicate AI illiteracy is in high demand. In addition, the reliance on ML models in the medical domains is undeniably beneficial, as in computer-aided diagnosis. Such motives inspired us to guide Senior Design students through the gateway of Machine Learning (ML) principles and applications in the light of both ground and cloud computations. First, the students would focus on building local native software, and, indeed, this would not suffice the need for high-end large-scale software.

In the context of healthcare, the integration of AI applications into cloud environments offers significant potential for innovation. AI-driven solutions can improve healthcare delivery to have better diagnostic accuracy and personalized treatment plans. Nevertheless, the healthcare domain presents unique challenges for new graduates, who must navigate complex regulatory environments, high data security, and privacy standards. To address these challenges and equip the next generation of engineers with adequate skills, we used a paradigm that focuses on clarifying the challenges and potential solutions for the designated healthcare problems.

Our work is structured into multiple stages, beginning with foundational concepts in machine learning, healthcare domain knowledge, and cloud computing. Students are then tasked with solving a real-world healthcare problem using AI techniques, culminating in benchmarking leading cloud platforms, Google Cloud Platform (GCP) and Amazon Web Services (AWS) with regard to switching to cloud healthcare systems. This paper documents and analyzes each stage's outcomes, highlighting our approach's effectiveness in preparing students to tackle the challenges of cloud-based AI applications in healthcare.

A. *Problem Definition*

At present, computer engineering students are expected to have adequate knowledge of cloud computation and AI basic techniques. Nevertheless, numerous students are not well familiar with the fundamentals of modern AI in general and cloud-based software production in particular. In this work, we aim to provide preliminary yet comprehensive guidance on generally learning AI and ML techniques and integrating them with cloud computing, with a particular focus on benchmarking different cloud platforms of Google Cloud Platform (GCP) and Amazon Web Services (AWS) in the light of AI healthcare applications.

II. Related Work

Artificial Intelligence (AI) and Machine Learning (ML) have become transformative forces across various sectors, particularly in healthcare, where they are revolutionizing diagnostics, patient management, and medical research. The integration of AI into healthcare systems promises not only enhanced accuracy and efficiency but also personalized care that was previously unattainable.

The role of cloud computing in enabling these advancements cannot be overstated. Cloud computing provides the scalable, flexible, and cost-effective infrastructure necessary to support AI-driven healthcare applications. As Armbrust et al. [1] discuss, cloud computing has emerged as a pivotal technology, offering on-demand access to computational resources and storage, thereby democratizing access to powerful IT capabilities. This has allowed healthcare institutions of all sizes to leverage advanced AI tools without the need for substantial upfront investments in hardware and software.

Xing, Giger, and Min [2] provide a comprehensive review of AI's trajectory in healthcare, highlighting how AI has evolved from rudimentary applications in medical image analysis to sophisticated tools capable of performing complex diagnostic tasks with a level of accuracy that often surpasses human experts. Their work underscores the significant strides AI has made in radiology, where machine learning algorithms are now routinely used to detect and classify medical conditions in imaging data, leading to earlier and more accurate diagnoses. The scalability offered by cloud computing, as noted by Armbrust et al. [1], has been instrumental in processing the vast amounts of imaging data required for these AI algorithms to function effectively.

Flores-Alonso et al. [3] emphasize the importance of integrating AI education into undergraduate engineering programs. Their study illustrates how early exposure to AI concepts prepares future engineers to tackle the challenges of modern technology-driven industries. By incorporating AI-related content into the curriculum, students not only gain theoretical knowledge but also develop practical skills that are essential for innovation in fields such as healthcare. The approach advocated by Flores-Alonso et al. highlights the critical role education plays in equipping the next generation of engineers with the tools needed to leverage AI effectively in their careers.

The ISIC (International Skin Imaging Collaboration) dataset has become a critical benchmark for developing AI models for skin lesion analysis, particularly in detecting and classifying melanoma. Studies leveraging this dataset have achieved significant advancements in diagnostic accuracy for dermatology. However, the dataset's complexity, including challenges like class imbalance and variability in image quality, necessitates robust data preprocessing and standardized evaluation metrics. These benchmarks ensure consistent model assessment, facilitating meaningful comparisons and driving progress in AI-driven healthcare [7]–[10].

Beyond dermatology, AI's role in other areas of personalized medicine is equally significant. Collins and Varmus [5] emphasize that precision medicine, which tailors treatment plans to individual patient profiles, is greatly enhanced by AI's ability to analyze vast datasets and predict treatment responses. Cloud computing, with its ability to store and process massive amounts of genomic and patient data, plays a critical role in making these AI-driven personalized treatments feasible on a large scale.

The drug discovery process, traditionally a lengthy and costly endeavor, has been revolutionized by AI. Kermany et al. [8] demonstrated how deep learning models could be employed to identify medical diagnoses and treatable diseases based on image data, thus accelerating the drug discovery process. The cloud's ability to provide virtually unlimited computational resources, as described by Armbrust et al. [1], enables researchers to run complex simulations and data analyses that were previously impractical, significantly speeding up the development of new treatments.

AI's integration into patient management systems is another area where significant advancements have been made. Bates et al. [6] discuss the use of AI in managing chronic diseases, where predictive algorithms monitor patients in real-time, anticipate complications, and suggest interventions. Gulshan et al. [11] further support this by demonstrating the effectiveness of deep-learning algorithms in detecting diabetic retinopathy, showcasing the potential of AI to improve patient outcomes in chronic disease management. Cloud computing supports these AI applications by enabling real-time data processing and seamless integration across different healthcare systems, ensuring that patient data is always up-to-date and accessible.

While the benefits of AI in healthcare are clear, Xing et al. [2] also highlight the ethical and regulatory challenges that accompany these advancements. Issues such as data privacy, algorithmic transparency, and the potential for bias in AI decision-making are critical concerns that need to be addressed. Armbrust et al. [1] also note that cloud computing introduces its own set of challenges, particularly regarding data security and regulatory compliance. Ensuring that cloud-based AI systems adhere to strict privacy standards and regulations is essential for their widespread adoption in healthcare.

Ching et al. [12] provide a comprehensive overview of the opportunities and challenges associated with deep learning in biology and medicine, highlighting the potential for AI to revolutionize these fields while also addressing the significant obstacles that must be overcome. Their work underscores the importance of collaboration between healthcare providers, researchers, and policymakers to navigate the ethical and regulatory challenges associated with AI, ensuring that these technologies are implemented responsibly and equitably.

This literature review highlights the critical role AI and cloud computing play in transforming healthcare, from improving diagnostic accuracy and personalizing treatments to enhancing patient management and accelerating drug discovery. The inclusion of benchmark datasets such as ISIC in this transformation underscores the importance of standardized, high-quality data in developing reliable and effective AI models. However, as these technologies continue to evolve, it will be essential for healthcare providers, researchers, and policymakers to collaborate closely to navigate the ethical and regulatory challenges, ensuring that AI-driven innovations are implemented responsibly and equitably.

III. Approach

This section outlines the approach employed in the study, emphasizing the pedagogical approach, the design process, the benchmarking process, and the required deliverables from the student teams. These outlines are covered in five chronological stages.

A. AI Literacy

We executed the pedagogical process by initially teaching multiple fundamental AI concepts to leverage the students' qualifications and understanding in both AI and medical domains. Through biweekly meetings, which included 30-minute instructional sessions and technical support sessions, we managed to build up adequate knowledge and hone essential analytical and developmental skills required for general healthcare AI applications.

1) AI Fundamentals: Initially, we introduced the students to the fundamentals of ML techniques, including supervised learning, unsupervised learning, and transfer learning, which enables the use of pre-trained models to accelerate the development of new applications. This foundational knowledge is a crucial prerequisite for understanding how AI models can be trained, optimized, and deployed effectively to address real-world challenges.

In addition, we delivered the concepts behind the importance and the role of AI in healthcare in general, cancer and allergy prognosis and diagnosis in particular, while emphasizing the prominent limitations of the medical domain, like the unavailability of sufficient data to build and deploy robust models. Subsequently, we gave the students the basic tools to crack through the challenge of AI-driven healthcare services.

2) ML Platforms: We provided an overview of machine learning frameworks like Keras and TensorFlow, demonstrating how to start and use both AI and ML projects. We began by introducing the core concepts of Keras as a high-level API for building and training neural networks and TensorFlow as a comprehensive end-to-end open-source platform for machine learning. A crucial milestone was to demonstrate the processes of building and training different models. Students were instructed to experience various model architectures as a starting point.

3) Transfer Learning: The next milestone was to illustrate how to utilize pre-trained models. Loading, training, and saving models and model weights were essentially delivered concepts. Indeed, transfer learning is a vital technique to increase accuracy, robustness, and efficiency in the light of ML.

4) Traditional Engineering vs. AI: In engineering, traditional programming and machine learning can be used disjointly to reach similar outcomes given the appropriate application. Nonetheless, for many applications and challenges, engineers have to completely define the boundaries separating those approaches apart, as some challenges are more fitting for one and not the other at the same time. For instance, medical imaging has undergone significant changes transitioning from traditional approaches to machine learning, as it was often proven hard to produce a reliable program of imaging using only traditional techniques.

For a student to have a simple distinction between what they are used to using in undergraduate study and AI, we instructed the students to attempt to solve the assigned challenges without relying on machine learning. Additionally, they reported the results of this attempt and their respective analysis for why they might have failed to achieve the same results in both scenarios. Normally, students have built up some experience of how to build some programs trying to conquer some challenges. However, if the givens are different, they cannot achieve the same results. In ML, it is the program, rather than the solution to the challenge itself, that the model seeks to discover, causing significant confusion among students.

5) Cloud Platforms: We introduced an overview of platforms like AWS and GCP and how to start and use both clouds. we began by demonstrating the two clouds, their core services, and how they can be accessible as a computational asset via the internet. We introduced EC2/Compute Engine for computing and S3/Cloud Storage for object storage. To solidify understanding, students were tasked with creating basic resources on both platforms and conducting an initial comparison between them. Emphasizing the pay-per-use model, we encouraged cost-conscious resource management.

B. Medical Domain Knowledge

As limited as the students' experience in the medical background can be, we needed to introduce them to the general background regarding the medical domain knowledge as an interdisciplinary field with ML. Artificial Intelligence applications in the medical domain face distinct problems, including general medical terminology and expected challenges, as well as application-specific obstacles, since in order to reliably address the healthcare ML application challenges, we have to cover and understand the different aspects of the whole medical process. The included domains revolved around focusing on cancer and allergy diagnosis and prognosis in some cases.

1) Healthcare Challenges: Basically, we had three fundamental problems; first, to address the challenge of detecting skin cancer given visually labeled data. Second, to address the challenge of breast cancer diagnosis and prognosis using spectral FTIR-extract data; third, to address the challenge of allergic reactions given personal records and allergens. This milestone included introducing and explaining each of these projects, challenges, the datasets and labels included, and the potential solutions for them. We emphasized understanding specifically the various datasets for the projects as inputs and outputs to find the correlation between them. In other words, to better explain and plan a solution for each challenge we formulated the problem as specific as possible in a technical manner.

C. Metrics and Measurements

In order to accurately and efficiently benchmark and compare systems, sets of rules have to be defined and standardized to be able to critically assess the systems and reliably decide which one is better. Thus, we made sure that the students grasped these concepts and implemented them during the process of development, both ground and cloud.

Metrics are the attributes taken to assess quantifiable indicators, such as the performance or quality of a system, process, or component. In the context of AI applications, metrics can be set in two ways: first, to assess the model's accuracy, precision, recall, or error and to assess the performance of the whole system running, including the processing time, scalability, or pricing. Metrics are, therefore, a standardized way of assessing and comparing effectiveness between different systems or platforms.

Measurements, for each object, are real data or values of measurements gathered during the measurement process of a system or component using the metrics defined. For example, if response time was the metric to measure, the measurement is the specific time read during the operation of the system. Measurements provide concrete data, which can be used in analyzing performance against a set of metrics.

Criteria are rules used to judge and ultimately make decisions about a system or a process. For example, criteria could be the requirements the platform should meet: cost-effective, scalable, compliant with healthcare security requirements, and integrated with other systems for medical record tracking. Criteria form the basis for choosing the most suitable platform by ensuring that all essential factors are considered.

While metrics and measurements are closely related, metrics are the broader categories or aspects of performance being evaluated, and measurements are the actual data points collected. Criteria, on the other hand, are the guidelines that determine which metrics and measurements are most important in the context of a particular decision or comparison. Metrics and measurements are more focused on data and performance, whereas criteria are about aligning these data points with specific goals or standards.

To effectively choose the best cloud platform for a given application, it is important to initially define the decisive metrics that align with the goals of the project. For example, in a healthcare setting, key metrics to be measured might include:

- Response Time: How quickly the AI application processes and returns results.
- Accuracy: The correctness of the AI model in predicting health conditions.
- Scalability: The ability of the platform to scale up amounts of data or users with steady performance.
- Security Compliance: The platform's adherence to healthcare regulations like the General Data Protection Regulation (GDPR).
- Cost Efficiency: The overall cost associated with running applications, like storage, computational resources, and operational expenses.

Once these metrics are defined, measurements can be taken by deploying the AI application on different cloud platforms and recording data relevant to each metric. For example, the response time of the AI application can be measured in milliseconds on both AWS and GCP, while accuracy rates can be assessed by comparing the outcomes of medical diagnostics performed on each platform.

Criteria then serve as guidelines in carrying out decision-making. For instance, where the primary consideration is with respect to patient data security, the most critical criteria may well be security compliance, and those platforms not meeting required standards would be excluded from consideration regardless of performance on other metrics. If cost efficiency is a major concern, the total costs of operation measured against the performance on each platform would be weighed to determine where to find the best value.

Once the criteria have been established and metrics are measured, benchmarking becomes a crucial step in the process. Benchmarking involves comparing the performance metrics of different systems or platforms against each other or against a predefined standard. This process enables students, researchers, and professionals to objectively assess which platform or system best meets the defined criteria. By systematically comparing measurements—such as response times, accuracy, scalability, and cost efficiency—across different platforms, benchmarking helps to identify strengths and weaknesses and provide practical insights. Oftentimes, in order to draw a decisive conclusion, a holistic quantification from the comparison of measurements is implemented. Afterward, the decision-maker would be capable of mathematically specifying the weights for each decision, based on the defined criteria. For instance, binary quantification can be a reasonable approach to how to weigh each measurement for each metric, and the option that outweighs the other would be the best to take.

In the context of AI applications in the domain of healthcare, benchmarking can reveal which cloud platform offers the best balance of performance, cost, and compliance, ensuring that decisions are made based on comprehensive and data-driven evaluations. This process is essential for making informed decisions systematically and plays a critical role in validating the effectiveness of AI models and systems in real-world scenarios.

The solution framework we propose to address the challenges is generally based on two phases, each designed to foster the experience of building both local and cloud-based AI systems. After the students met the requirements based on the prior pedagogical stages, we started executing the first phase of development.

D. Solution Development

1) Via Ground Computation: At this point, students should have built up adequate knowledge and experience to address the mentioned healthcare challenges. As mentioned earlier, they were instructed to follow two phases of how to conduct the solution.

The first phase aimed for the students to understand and build a healthcare system locally in a ground computing environment. Then, they are required to analyze and benchmark the capabilities of the produced software, which will give the students insight into the requirements for successful medical software. As a result, students should be aware of the relationship between the hardware capabilities, software design, and financial aspects of AI software production.

2) Via Cloud Computation: The second phase aimed to experiment and compare two leading cloud platforms, Amazon Web Services (AWS) and Google Cloud Platform (GCP). Basically, by using the designated cloud platforms, multiple essential steps were to be followed:-

1) Running the services on both clouds.
2) Building and training the same model on both clouds.
3) Defining the required metrics and measurements.
4) Defining the criteria, both qualitatively and quantitatively.

Through senior design projects, students conducted detailed benchmarking of these platforms by analyzing various aspects, like cost efficiency, scalability, and performance. Subsequently, students were instructed to design and deploy scalable, efficient, and reliable AI systems for healthcare solutions since, in this phase, students are required to migrate their designed system from ground computing to cloud computing. Subsequently, this migration proposes other considerations and complications regarding the hardware and the nature of the services offered by the clouds. They need to define evaluation criteria for a valid comparison between the two environments. For instance, testing the computational power of each environment would be a crucial criterion to consider, specifically in the medical domain.

The two phases are designed to be incremental and chronological, providing students with a comprehensive understanding of AI and cloud computing in healthcare applications.

E. Senior Design Projects

We recruited 36 students, distributed over 6 groups. Each group is assigned to a senior design project for an AI healthcare-related challenge, which ensures a diverse range of perspectives and approaches to the study. The 6 groups included 2 groups addressing the skin cancer diagnosis challenge, 2 groups addressing the allergy diagnosis challenge, and the last 2 included the breast cancer post-operation prognosis challenge. Indeed, all the groups were instructed to submit the same deliverables by the end of the project, which included developing two analogous healthcare systems, one deployed on local computational environments (ground) and the other on cloud platforms (AWS and GCP). Indeed, the results of deployed systems are summarized and analyzed by students to evaluate and conclude the whole educational process.

1) Projects Description: The skin cancer project includes an open-source skin cancer dataset that includes thousands of labeled skin cancer images. The labels can be categorized in various ways, like diagnostic, clinical, or technological attributes. Diagnostic attributes include the basic classification of whether it is benign, intermediate, or malignant, while the clinical includes other information about the subjects themselves, like gender and personal history of melanoma, and technological attributes include dermoscopic type and image type.

Second, the pathology project includes hard-to-acquire, uncommon data about treated breast cancer subjects. The samples are originally collected by acquiring blood serum from biopsy, and performing FTIR analysis, then encoding the results. clinical attributes and results from medical prognosis are included with the dataset, which serves the purpose of prognosis through AI techniques.

Third, the allergy project includes data about allergies, products, and product components, including allergens. The data serve the purpose of building up a system fully capable of detecting the allergies based on previously observed allergies and the query product. In other words, the project aims to classify and detect penitential allergies and allergic reactions based on given merchandise and known allergies.

2) Benchmarking: As mentioned, the metrics and criteria used for the comparison and benchmarking are some of the most crucial points throughout the process, as the accurate assessment of the systems would convey effective insight into comparing and ranking different environments running the same models.

Indeed, our comparison of cloud services incorporated both quantitative and qualitative metrics. While quantitative metrics provided a straightforward basis for comparison, qualitative factors such as environment setup and ease of use also played a crucial role in the assessment. To standardize these qualitative measures, we converted them into quantitative scores, allowing for a more balanced and comprehensive evaluation. For example, we quantified the teams' feedback on their experiences and satisfaction by translating linguistic feedback into numerical values. This approach ensured that qualitative insights were given comparable weight alongside traditional quantitative metrics, enabling a more holistic comparison of the cloud platforms.

Undoubtedly, the results from a complete comparison between cloud services and local or personal services were quite determined beforehand since local deployment does not match the determined task achievement; it is only deployed on one device with less competence compared to the cloud. Thus, it is used as a proof of concept and to tune the ML models as much as possible. As illustrated in Table I, the metrics mainly included Performance, Effectiveness, Scalability, and Cost.

Metrics	Description	Evaluation Method
Performance	Response time, throughput	System monitoring and logging
Effectiveness	How effective are AI Models	Models Accuracy and Loss values
Scalability	Ability to handle increased loads	Load testing and scalability tests
Cost	Initial and ongoing costs	Financial analysis

TABLE I: Quantitative Metrics

Generally, the designated criteria cover essential metrics to be satisfied in a platform, or the platform would fail inevitably to do the task, like security measures. For a medical application, in particular, low-security standards would lead to disasters. Overall, both clouds meet the basic security requirements. Therefore, students focused on the disparities. Each team designed and developed two complete cloud-based healthcare systems, including system architecture, setup environment, system User Interface (UI), and system back-end. The healthcare AI applications should be identical on both AWS and GCP, then systematically record quantitative measurements and qualitative observations before and after the deployment.

3) Deliverables and Follow Up: The deliverables included documenting the work with a report and a poster for representation of the findings, highlights, and lessons learned. The report included the technical progress along with the results, including the evaluation metrics and measurements and the designed and produced software. Team contribution matrices, which state each student's role and work, were monitored and analyzed throughout the whole duration of the project. Therefore, we would rely on those results and deliverables from students, in addition to following them up throughout the project, in order to be able to assess how engaging, impactful, and effective our approach was. Students included their repositories of code for both systems to trace and reproduce their work.

4) Evaluation Process: The evaluation process is a critical checkpoint that we relied on to assess how successful our approach was throughout the various stages of the project. In general, we relied on a qualitative approach to analyze the outcomes and goals achievement of the projects. First, we relied on the bi-weekly meetings to monitor and analyze the progress of the groups in the mentioned projects. Second, we used the progress documentation and students contribution matrices to show the overall progress of the team holistically and individually to address any singularities hindering the team's progress. Third, we relied on assessing the students' deliverables, including the detailed results, deployed system, and documentation, to conclude how far the students went and how deep and broad their knowledge had become before and after the work.

By following this structured approach, students gained a holistic understanding of AI and cloud computing in healthcare, from foundational concepts to practical deployment and benchmarking. This approach ensures that they are well-prepared to navigate the complexities of modern healthcare technology and make informed decisions about cloud-based solutions.

IV. Results and Discussion

Although students followed the same scheme for benchmarking, they had some disparities of how they would benchmark and compare both of the systems deployed on AWS and GCP. The projects share the same goals, but they basically have different projects with different specifications and technical requirements. Different teams relied on different machine instances from AWS, GCP, and Google Colab.

In this section, we illustrate how each system performed in both quantitative and qualitative metrics. On average, the results indicate that both clouds are competitively close regarding performance and, in some cases, the cost. However, when it comes to building up environments and convenience, GCP has the upper hand by a minor distance.

Results are summarized in two points. First, the AI technical results represent how well the produced models are performing. Second, the process of comparing and benchmarking the GCP and AWS clouds.

A. *Skin Cancer Projects*

1) Team 1: **AI Results:** Figure 1 illustrates the accuracy of the models built by the team in their attempt to tackle the challenge of skin cancer. The team achieved an accuracy of 94%, with a sensitivity of 90% in detecting malignant skin lesions.

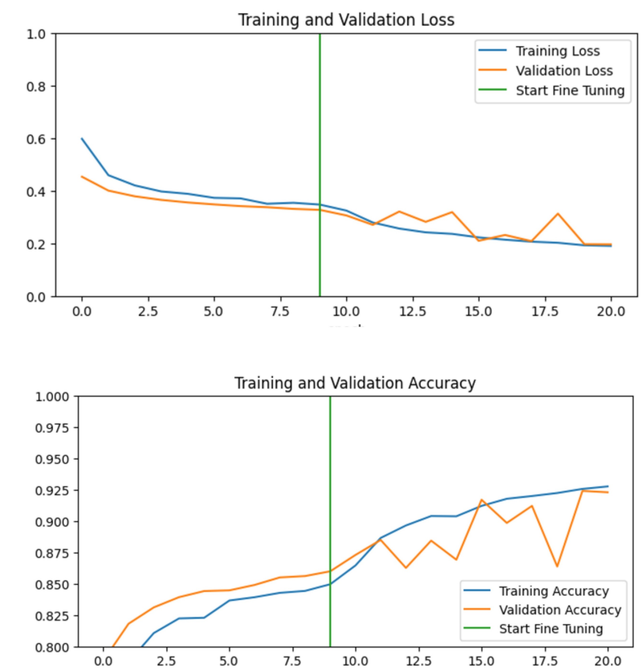

Fig. 1: Team 1 - Skin Cancer Project - Model Loss and Accuracy

Cloud Benchmarking: Based on this group's work, AWS offers moderate cost efficiency. Yet, it is outperformed by GCP for the overall performance. Generally, AWS implements a detailed architecture with various services, making it robust but complex, while GCP's streamlined approach results in a simpler and more efficient deployment process. Configuration and resource management in AWS are more complex due to multiple services and detailed setup requirements, while GCP provides a simpler, more user-friendly process. Moreover, in terms of security, AWS relies on pre-signed URLs for secure data transfer and automatically deletes images post-processing. On the other hand, GCP employs IAM permissions and secure storage in Cloud Storage. Overall, GCP offers more efficient costs and improved performance - see Table II. GCP additionally offers a simpler deployment process with ease of use, making it a more preferred option. In the attempt to rate the user experience both systems, GCP would receive 4 points and AWS 2.5 out of total 5 points. The training time in Table II refers to the training time before and after the fine-tuning. The AWS instance used is Amazon EC2 t2.medium, and the GCP is Google Compute Engine e2-medium.

Metric	AWS	GCP
vCPUs	2	2
RAM	4.0 GB	4.0 GB
Training (before tuning)	2511.56 secs	3035.61 secs
Training (after tuning)	5079.30 seca	4226.84 secs
Total Training Time	7590.86 secs	7262.45 secs
Evaluation Time	125.09 secs	33.44 secs
CPU Utilization	Lower	Higher

TABLE II: Team 1 - Model Training Performance Comparison between AWS and GCP

2) Team 2: **AI Results:** Figure 2 shows the designed software for the skin cancer detection. Generally, the implemented models showed a high testing accuracy with a value of approximately 98%.

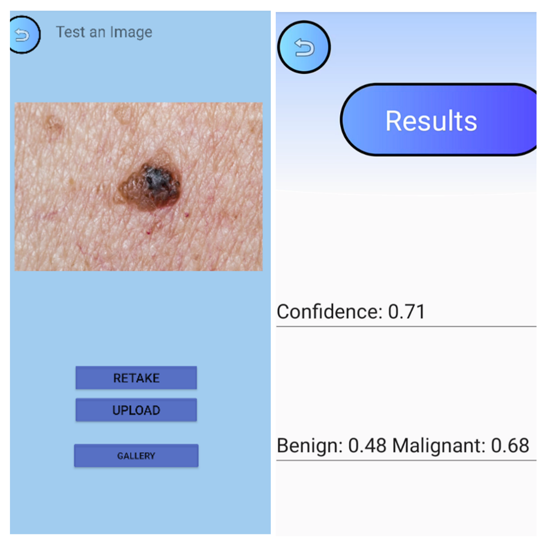

Fig. 2: Team 2 - Skin Cancer Project - Designed Software for Cancer Detection

Cloud Benchmarking: For the comparative analysis between AWS, GCP, and ground computing environment, in terms of training accuracy, they achieved close results, around an accuracy of 98.28%. Yet, for the validation, GCP outperformed AWS. Similarly, the training losses were not afar between GCP and AWS. For the pricing, AWS was slightly cheaper than GCP, at $49.80, compared to GCP's $57.50. Over a span of a month, such results were estimated to cost around $240 for AWS and $284 for GCP, given the rate of $0.33 for the former and $0.39 for the latter. The documented results of accuracies and training losses are inclusive of 5 iterations for each environment, with a batch of 1 - see Table III For detailed computational performance, Table IV lists a detailed comparison based on the dataset generation and loading, preprocessing, model building, model training, model exportation, and the total time per iteration. Overall, the comparison indicates that GCP outperformed AWS in every aspect.

TABLE III: Team 2 - Performance Comparison for AWS VS GCP

	AWS	GCP	Local
Tr. Accuracy	98.28%	98.28%	98.28%
Val. Accuracy	49.10%	77.19%	98.0%
Tr. Loss	0.984	0.0991	0.0854
Val. Loss	1692.4597	6.5516	0.6469

TABLE IV: Team 2 - Time Comparison for AWS and GCP

Activity	AWS Time	GCP Time
Dataset Generation	4.6235 Seconds	0.991 Seconds
Dataset PreProcessing	0.5168 Seconds	0.4031 Seconds
Model Building	2.0074 Seconds	1.1855 Seconds
Model Training	60.044 Hours	50.850 Hours
Model Saving	0.9682 Seconds	0.8918 Seconds
Hours per Iteration	12.0088 Hours	10.1701 Hours

B. Pathology

1) Team 3: **AI Results:** Students achieved acceptable results given the proposed challenge in this project. They reached errors as low as 50 months of survival prediction after using normalization, compartmentalization, and encoding the data, which is relatively close to the data standard deviation. The setup they followed allows for the rapid analysis of complex pathology data, utilizing cloud computing to handle the processing load and delivering predictions back to the user in a user-friendly format. This architecture supports scalable, real-time diagnostics that can improve the speed and accuracy of pathology assessments.

Cloud Benchmarking: As illustrated in Table V, the difference between clouds in performance is negligible. The difference in accuracy is not vast, especially for this documentation, as it was recorded that the loss differs given the consideration of the random start. Thus, they were close. Students relied on Google Colab to train models, which was free to use. However, they have to use GCP and AWS for deployment, which requires monthly payments, as referred to in the table. Students on this team concluded that the differences between GCP and AWS are not vast, given their approach to what machines they specifically needed and their financial capabilities. Thus, it was a close call between the two clouds, considering the rough 3 points of rating for the qualitative metrics assigned to each cloud.

2) Team 4: **AI Results:** Figure 3 illustrates the designed system architecture. The students leveraged Keras and Random Forest to improve the breast cancer outcomes predictions, including cancer grade, cancer stage, and recurrence, achieving a precision rate of over 85% for the grade, 80% for the stage, and 89% for the recurrence. Figure 4 illustrates the accuracy

Criteria	**AWS**	**GCP**
Training Time	50 Seconds	51 seconds
Mean Squared Error	2651	2773
Machine Instance	ml.t3.medium	Vertex AI
Cost for Deployment	• \$0.05/hr • S3 storage: First 50 TB at \$0.023/GB	• \$0.05/hr • 0-2M calls: \$0 • 2M-1B calls: \$3.00 • 1B+ calls: \$1.50
Free Trial	Free tier for 12 months	\$300 free for 90 days
Deployment	Sagemaker with TF	Vertex AI with TF

TABLE V: Team 3 - Comparison between AWS and GCP

of the model training process and corresponding training and validation loss values.

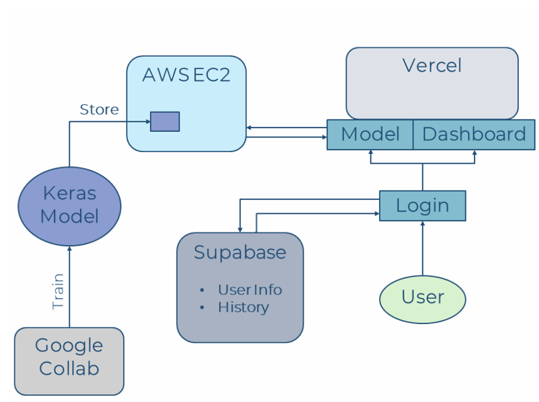

Fig. 3: Team 4 - Pathology Project - System Architecture

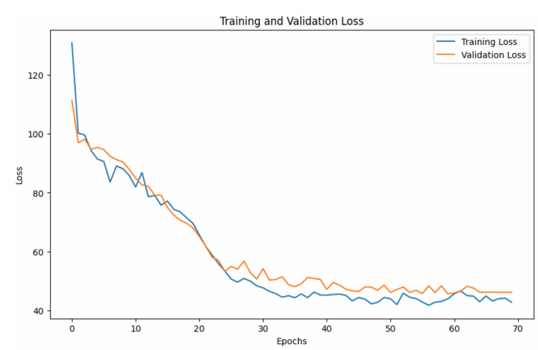

Fig. 4: Team 4 - Pathology Project - System Training

Cloud Benchmarking: Team 10 relied on the AWS EC2 instance. Initially, they were relying on AWS Lambda and Amplify, then they used EC2 and Vercel to successfully deploy their system. When it comes to accuracy, they scored a significant improvement in how to tackle the challenge of prognosis, as in this stage of the survival period, prediction, they managed to reach 37 months of an error, which is significantly high taking into consideration the difficulties prevailed in the training processes and data biases.

C. Allergy

1) Team 5: **AI Results:** For the allergy project, team 5 achieved acceptable results regarding predicting allergic reactions to some merchandise and skin care products, with a recall of %86 and F1-Score of %65 given 9 classes for classification.

Cloud Benchmarking: For hosting, the students relied on storing the application file as static content within the cloud object storage of both providers. On the AWS side, they stored the react app file within an S3 bucket, while GCP took an extra step, as they needed to provide a *yaml* file. In general, the GCP was preferred due to its convenience and ease of use for development and hosting compared to AWS. The performance difference was only marginal, and the differences in costs were not drastically vast - see Table VI.

Metric	GCP	AWS
Cost	$0.02	$0.00 (Free Tier)
Memory Usage	$>$ 512MB	700MB
Cold Start Time	11 seconds	10 seconds
Warm Start Time	0.6 seconds	0.4 seconds

TABLE VI: Team 5 - Comparison between GCP and AWS

2) Team 6: **AI Results:** Team 6 managed to score high results with an accuracy of 71%. They managed to furtherly push those results.

Cloud Benchmarking: Regarding the clouds, the team was familiar with AWS beforehand, which gave AWS a level of preference over GCP. The free trial plans were more suitable for the development over AWS, as they included 3-6$ per week in comparison with 20$ per week for GCP. In the end. they relied on qualitative approaches to rely on the deployment with AWS more than GCP.

V. Limitations

One of the biggest limitations was the difficulty of acquiring medical datasets. The pathology team had his hands over exclusive data for which they had to sign an NDA since new medical data and clinical information are extremely sensitive. In addition, they needed more time to analyze and have insight into the data itself to decide what ML approach they would need to follow, as the data could be used in the light of both supervised learning techniques of classification and regression and unsupervised learning. In addition, the data showed a high level of bias, especially due to its small size. Therefore, it was hard to tune the model to get better results without overfitting the models. The same scenario of lack of data applies to the allergy project. The skin cancer project required higher diversity regarding the taxonomy of cancer in the data as well.

VI. Conclusion

In conclusion, learning AI and ML principles is crucial for undergraduate teaching. Proposing real-world solutions, especially in healthcare, with the power of AI as a modern and vital stage that would prepare the students and empower them to be strong and skillful engineers. As engineers, they have to experience the power of cloud computation, understand how it works, how to reliably benchmark them, and the advantages and disadvantages of using either of the cloud platforms mentioned. AWS and GCP are some of the most famous cloud platforms, known for their good plans, power, and ease of use. Overall, in this work, GCP was proved to have a slight preference over AWS based on both quantitative and qualitative criteria. In the future, we aim to experiment with more advanced projects through which the students can gain more knowledge and field experience and, subsequently, have a bright AI future.

Acknowledgment

We like to express our sincere gratitude to the senior design students whose hard work, creativity, and commitment were key to applying AI to real-world healthcare challenges. Their innovative efforts have laid a strong foundation for future explorations in healthcare AI.

References

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). *A View of Cloud Computing*. Communications of the ACM.

[2] Xing, L., Giger, M. L., & Min, J. K. (2020). *Artificial Intelligence in Healthcare: Past, Present, and Future*. Journal of the American College of Radiology.

[3] S. Isaac Flores-Alonso et al., "Introduction to AI in Undergraduate Engineering Education," 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 2023, pp. 1-4.

[4] A. F. Salazar-Gomez, A. Bagiati, N. Minicucci, K. D. Kennedy, X. Du and C. Breazeal, "Designing and implementing an AI education program for learners with diverse background at scale," 2022 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, 2022, pp. 1-8.

[5] F. S. Collins and H. Varmus, "A new initiative on precision medicine," The New England Journal of Medicine, 2015.

[6] D. W. Bates et al., "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, 2003.

[7] Beam, A.L., & Kohane, I.S. (2018). "Big Data and Machine Learning in Health Care." *JAMA*, 319(13), 1317-1318.

[8] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., & Dong, J. (2018). "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." *Cell*, 172(5), 1122-1131.e9.

[9] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J.T. (2017). "Deep learning for healthcare: review, opportunities and challenges." *Briefings in Bioinformatics*, 19(6), 1236-1246.

[10] Topol, E.J. (2019). "High-performance medicine: the convergence of human and artificial intelligence." *Nature Medicine*, 25(1), 44-56.

[11] Gulshan, V., Rajan, R.P., Widner, K., Wu, D., Wubbels, P., Rhodes, T., Whitehouse, K., Coram, M., Corrado, G., Ramasamy, K., Raman, R., Peng, L., & Webster, D.R. (2016). "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India." *JAMA Ophthalmology*, 134(10), 1087-1094.

[12] Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M., Xie, W., Rosen, G.L., Lengerich, B.J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A.E., Shrikumar, A., Xu, J., Cofer, E.M., Lavender, C.A., Turaga, S.C., Alexandari, A.M., Lu, Z., Harris, D.J., DeCaprio, D., Qi, Y., Kundaje, A., Peng, Y., Wiley, L.K., Segler, M.H., Boca, S.M., Swamidass, S.J., Huang, A., Gitter, A., & Greene, C.S. (2018). "Opportunities and obstacles for deep learning in biology and medicine." *Journal of the Royal Society Interface*, 15(141).