# Multitask Learning or Transfer Learning? Application to Cancer Detection

Stephen Obonyo and Daniel Ruiru

*Faculty of Information Technology, Strathmore University, Ole Sangale Link Road, Nairobi, Kenya*

Keywords: Multitask Learning, Transfer Learning, Deep Learning, Cancer Detection, Medical Imaging.

Abstract: Multitask Learning (MTL) and Transfer Learning (TL) are two key Machine Learning (ML) approaches which have been widely adopted to improve model's performance. In Deep Learning (DL) context, these two learning methods have contributed to competitive results in various areas of application even if the size of dataset is relatively small. While MTL involves learning from a key task and other auxiliary tasks simultaneously and sharing signals among them, TL focuses on the transfer of knowledge from already existing solution within the same domain. In this paper, we present MTL and TL based models and their application to Invasive Ductal Carcinoma (IDC) detection. During training, the key learning task in MTL was detection of IDC whereas skin and brain tumor were auxiliary tasks. On the other hand, the TL-based model was trained on skin cancer dataset and the learned representations transferred in order to detect IDC. The accuracy difference between MTL-based model and TL-based model on IDC detection was 8.6% on validation set and 9.37% on training set. On comparing the loss metric of the same models, a cross entropy of 0.18 and 0.08 was recorded on validation set and training set respectively.



## 1 INTRODUCTION

In this study, we seek to investigate the effectiveness of the application of Transfer Learning (TL) and Multitask Learning (MTL) in Invasive Ductal Carcinoma (IDC) cancer detection. Multitask Learning can be defined as a learning method where a model learns by not only focusing on a single task $T$ but other auxiliary tasks $T_0$, $T_1$... $T_k$ as well. The signals from the auxiliary tasks generally improves model's performance on key task (Ruder, 2017). This characterization can be attributed to the fact that MTL architecture allows for sharing of informative representations among the tasks involved. Further, MTL has been shown to improve the ability of the model to generalize well on unseen instances (Ruder, 2017).

In contrast, Transfer Learning (TL) focuses on the improvement of a model's performance via transfer of knowledge from an already existing solution (source task) often within the same domain. Given a problem (target task), the aim of TL is to improve the performance by combining source task knowledge representations and related data (Torrey and Shavlik, 2010).

MTL and TL are somewhat related but the flow of information between the two is restricted as shown in Figure 1. In TL, information is unidirectional i.e. the flow is from source (already learned representations) to the target (new problem). On the other hand, in MTL the information flow is unrestricted and information can flow in any direction (among all the task related models). This behaviour is captured by Figure 1.

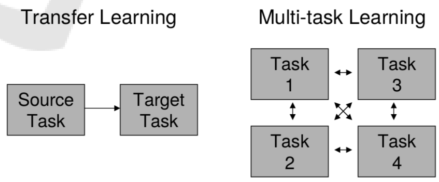


Figure 1: Multitask Learning vs Transfer learning. Source (Torrey and Shavlik, 2010).

This research compares the application of both MTL and TL to cancer detection. The tasks which were involved include detection of i) Invasive Ductal Carcinoma (a common sub-type of breast cancer), ii) skin cancer and iii) brain tumour. Based on these tasks both MTL-based and TL-based models were developed and their performance compared.

The key learning task for MTL-based model was IDC detection. Skin and brain cancer detection were auxiliary tasks. The TL-based model was developed by first training the base network to detect the skin



Obonyo, S. and Ruiru, D.
Multitask Learning or Transfer Learning? Application to Cancer Detection.
DOI: 10.5220/0008495805480555
In *Proceedings of the 11th International Joint Conference on Computational Intelligence (IJCCI 2019)*, pages 548-555
ISBN: 978-989-758-384-1

cancer. The learned representations from the model were transferred and used to detect IDC. The outcome of these two learning methodologies were presented and discussed on the results and discussion sections of this document.

Cancer research is likely to benefit from both MTL and TL learning approaches since they hold the potential towards developing robust models, even when relatively smaller datasets are available. This has translated to a development of stable models with an ability to generalize well on unseen examples.

# 2 RELATED WORK

## 2.1 Machine Learning in Cancer Medical Imaging

Cancer is by far the most adaptive and self-sustaining conditions currently known within the medical field. It dynamically interacts with its micro-environments to constantly thwart the efforts of practitioners, researchers and most importantly the affected patients.

Various complexities associated with the disease results in endless dilemmas at the different stages of its management, including the need for reliable early detection (Bi et al., 2019). Moreover, there are the problems associated with the accurate identification of preneoplastic and neoplastic lesions; the tracking of tumors; resistance to treatment; and the regulation of infiltrative tumor margins in surgical procedures etc. However, technological advances particularly in medical imaging and the identification of bio-markers hold great promise in addressing these challenges.

Machine Learning and its sub-branches have the ability to automate image analysis and this could potentially bridge the gap between cancer medical imaging and the accurate interpretation of conditions (Klang, 2018). Currently, there is disparity in manual cancer diagnosis. This hinders the treatment and the probable recovery of patients. These conventional or manual method of image evaluation rely heavily on the qualitative features of tumor cells such as density, pattern, cellular composition and anatomic relationship, among others. These features are hard to accurately determine due to the varying image dimensions.

In comparison, there exist radiomics which use quantitative features to analyze radiographic images. Broadly, radiomics uses the quantitative features of size, shape and textual patterns to describe medical images, which are better representations of tasks in ML (Bi et al., 2019). These types of descriptors could thus facilitate the role of Artificial Intelligence in medicine as the field has made great strides in automating the quantification of medical patterns. Deep learning in particular has the most promise having developed various models for learning and matching features in different problems. The ability of the implemented algorithms even surpasses those of human expert which further defines their significance in task-specific functions as they can be specialized as needed. Moreover, they are able to overcome the barriers of large data sets including the ability to withstand noise in foundation truth tables. In all, the capabilities of deep learning could provide exceptional insights into both qualitative and quantitative analysis further helping medical evaluations. This facilitation could for instance be achieved by the precise delineation of tumors, parallel tracking of lesions and the cross-referencing of tumors in related fields. Ultimately, deep learning methods promise the greatest generalization capabilities through the transfer of insights across various medical domains. It is these benefits among others that could provide the health care industry with the necessary tools for future breakthroughs.

## 2.2 Transfer Learning

The idea of generalizing models is an important element of transfer learning which in recent times has increased the popularity of the technique. Like in many other domains, transfer learning in medical imaging aims to transfer information from a particular classification problem(s) (source) to another (the target), hence improve the performance of the final classifier. In cancer treatments, this facilitation is important as the field has minimal data owing to the expenses involved and the few breakthroughs seen so far.

### 2.2.1 Transfer Learning Approaches

Despite the progress made in transfer learning there exist a lot of inconsistencies in the classification of its sub-branches. Traditionally, the categorization has been done using three main groups which are based on the similarities between domains and the availability of labeled/unlabeled data. Inductive, transductive and unsupervised transfer learning have thus been the three main categories. However, advances in deep learning have increased the scope of TL and today have led to a more flexible taxonomy having two main categories (Asgarian, 2019). This new taxonomy is based on the similarity of domains and has two major groupings: ***Homogenous transfer learning and heterogeneous transfer learning***.

#### 2.2.2 Homogeneous and Heterogeneous

In homogenous TL, both the source and target tasks have the same feature and label space. As such, the aim is usually to bridge the gap of data distribution that exist in the two instances (source and target). The somewhat reverse outcome is exhibited by heterogeneous TL as the source and target tasks have different feature spaces (non-overlapping). However, for the label space, a unique set up exists where the source and target domains can either share or have different data labels. From these broad classifications of TL, the solutions offered by the technique can be summarized into five different classes as outlined below.

1. ***Instance-based Approaches*:** Tries to re-weight samples in source tasks to correct marginal distribution differences. These re-weight instances are then directly applied in target domains during training
2. ***Feature-based Approaches*:** Are applied in both homogeneous and heterogeneous problems where in the latter setup they aim to reduce the differences in feature spaces. In homogeneous problems they are then used to correct the marginal and conditional distributions
3. ***Parameter-based Approaches*:** These approaches transfer knowledge using the shared parameters of the domains involved
4. ***Hybrid-based Approaches*:** These techniques transfer knowledge through both the instances and parameters shared by the tasks involved
5. ***Relational-based Approaches*:** These final approaches use the common relationships between the source and target domains to transfer knowledge (Asgarian, 2019).

#### 2.2.3 Empirical Results

Motivated by the current trends in deep learning, machine learning researchers have moved to develop algorithms that automatically classify cancer medical images. In particular, there has been a great emphasis on the transfer of features from pre-trained models due to the limited datasets (training corpus) found in the cancer domain. This borrowing of factors, better known as transfer learning has yielded better classification results and even helped generalize models. Take the example of transfer learning in the early gastric cancer classification as done by Liu et al. (2018). Using Magnifying Narrow-Band Imaging Images (M-NBI), this group of researchers were able to attain higher prediction accuracy with TL as compared to traditional ML methods. On average, a 96 percent accuracy was achieved, a value that on occasion improved by either fine tuning the final layers or all the layers of the applied model (Convolutional neural networks (CNN)). To further test the credibility of the result, different variations of CNN were used namely; VGG-16, InceptionResNet-v2 and Inception-v3. Ultimately, the research conducted found that the performance gain increased as the convolutional layers were fine-tuned with natural data (Liu et al., 2018). Furthermore, the amount of the input data (images) influenced the final result of the deep learning models. Since, the field in question has limited data (cancer and more specifically M-NBI), transfer learning provided the means to meet this functional requirements and hence improved the models' performance.

#### 2.2.4 Theoretical Framework

TL is common in deep learning owing to the amount of data needed to train models. Deep learning models require lots of data to make any meaningful predictions which often is not available. TL therefore works because it enables networks to use features learned in previous tasks by mixing and matching their functions into new as well as meaningful combinations. It is the new collaboration that helps improve the classification of a model. This outcome is observed in both theory and practice as models converge faster and are more accurate with TL as compared to when they are randomly initialized. Therefore, TL not only improves prediction results but also helps to train models faster.

Its mathematical representation highlights its theoretical background. Defining a domain $D$ as a two element tuple consisting of a feature space x and probability $P(X)$ (while space $x = a$ sample data point), then Domain $D$ can be defined as

$$D = x, P(X). \tag{1}$$

Note: In probability

$$P(X), X = x_1, x_2, \ldots\ldots\ldots\ldots, x_n \tag{2}$$

i.e.

$$x_i \varepsilon X \tag{3}$$

Additionally, if $X_i$ is a specific vector. A task $T$ can be accurately defined by corresponding tuples of y as label space and n as the objective function. Therefore, for the given domain (D), Task $T$ can be represented as

$$T = y, P(Y|X) = y, nY = y_1, \ldots\ldots., y_n, y_i \varepsilon Y \tag{4}$$

## 2.3 Multitask Learning

Deep learning models use a combination of many hidden layers and parameters in their learning process to give results. As such, they require lots of data. Cancer like many other medical fields does not meet this data requirement more so, because it applies manual labor to label data instances (Zhang and Yang, 2017). It is therefore, a perfect case for applying multitask learning (MTL) where useful information from multiple relevant tasks are used to alleviate the problem of data sparsity. MTL has been a promising field in machine learning since its initial formulation by Caruana (1997). Broadly, the goal of MTL is to leverage useful data/information found in multiple learning tasks to get more accurate learners. Of course, this objective assumes that the tasks (or their subsets) are related. Empirically and theoretically, jointly learning various tasks has been found to give better performances than when learning is done independently. Moreover, based on the tasks, MTL can take different setup which outlines its effective classification as: MTL supervised learning, MTL unsupervised learning, MTL semi-supervised learning and MTL online learning among others.

MTL helps to promote the notion that machines can mimic human learning activities as people transfer knowledge from different tasks to further others. For instance, the skills of long jump and, running track and field can facilitate each other, hence improve the performance of an athlete. Thus, MTL is simply an inductive transfer mechanism that aims to improve the generalization of machine learning models (Caruana, 1997). A concept (generalization) that it fulfills by leveraging domain-specific data from related activities through parallel training. Therefore, the training power of the additional tasks acts as an inductive bias. In this case, an inductive bias hails from its general definition which is anything that influences an inductive learner to prefer certain hypotheses as compared to others.

### 2.3.1 Empirical Studies

Most of observation studies of MTL have focused on feature selection problems where some attributes in multi-source data have been used in classification of regression experiments. In most cases, the features in question have been related even though they are derived from different data sources. Based on these underlying relations, it has been found to be easier to jointly select the necessary attributes (features) from various sources using joint selection regularizers. These regularizers, which are simply select constraints, have been found to improve the performance of classification models as compared to other conventional techniques that evaluate features individually based on their data sources. Examples of regularizers commonly introduced include joint sparsity, graph sparse coding, graph self-representation and low rank. It is the inclusion of these elements that has helped MTL deal with complex worldly problems such as the diagnosis of neurodegenerative diseases (Bib, 2019). Using structural Magnetic Resonance Imaging (sMRI), researchers have been able to predict the values of various types of clinical scores in these conditions, including their specific subject diagnostic labels. An example of this success is highlighted by the study of Alzheimer disease (AD) where clinical scores such as Mini-Mental State Examination (MMSE) and Dementia Rating Scale (DRS) have been used to grade the healthiness (functionality) of the brain.

As specified by MTL principles, the classification in this instance is based on the prediction of a target output. Because the target outputs, such as diagnostic labels and clinical score, are related then one gets better results unlike when each task is learned independently. It is this 'similar' approach that has led to the recent success of self-driving automation systems. In this case, images from cameras attached to subjects are used to detect objects (road signs, traffic lights etc.) which are then fed into neural networks to train a model for autonomous driving. A more robust system is acquired because the model gets to learn multiple objects simultaneously.

### 2.3.2 Multitask Learning Approaches

From the discussion above, MTL is simply a type of inductive transfer which improves algorithms by adding an inductive bias. This bias helps a model discriminate some attributes and thus, prefer some hypotheses over others. $\ell_1$ regularization is the most common type of inductive bias known in ML and is often used to get preferences for various sparse solutions. In contrast, MTL attains its inductive bias through auxiliary tasks which through their contributions models certain hypotheses inclinations. To achieve its goals, MTL commonly employs two contrasting ways in deep neural networks. They are; hard and soft parameter sharing (Ruder, 2017) The shared element comes from segmentation (sharing) of hidden layers.

1. ***Hard Parameter Sharing*:** Its application in neural networks goes back to Caruana (1997). It shares the hidden layers between all tasks involved but also maintains a few task-specific output layers. Because of its efficiency and simplicity

it is the most common approach of MTL. It also reduces the risks of over-fitting, a result that stems from the ability to develop a model that represents all tasks and not just the original concept (task).

2. ***Soft Parameter Sharing*:** On the other hand, this approach sees each task having its own model plus their own set of parameters. To encourage a similarity between the distinct parameters, the distance between them in the overall model is regularized using a bias function say $\ell_2$. This application of inductive bias functions, explains the huge inspiration that regularization techniques have had on the constraints of soft parameter sharing, an outcome that still stands today.

#### 2.3.3 Empirical Results

As highlighted before, MTL helps with the simultaneous solving of multiple tasks by optimizing several loss functions instead of one. It is this application that has seen it applied in several fields such as cancer diagnostics. Khosravan and Bagci (2018) specifically applied the technique in lung cancer and eventually were able to overcome their limitations of labeled data for task segmentation.

Having the highest mortality rate among cancer affiliated deaths, lung cancer has invoked a lot of research in attempts to yield conclusive results. This interest has produced many systems and models but they all seem to suffer from the same problem of false positive results. Additionally, the limitations of data segmentation (initial step of data quantification) lower the performance of the developed model. Khosravan and Bagci (2018) improved on this available models by incorporating MTL into their 3D encoder-decoder CNN structure. In doing so, they shared underlying features of tasks and trained single models using shared features that are essential in lung cancer screening. Eventually, the finding of their study saw the importance of MTL in semi-supervised learning where improved results are obtained even without large data sets. Essentially, minimal labeled data is needed when features are shared between tasks as they get to learn from one another. Moreover, the final model was easily generalized not accounting for the reduced false positive result.

#### 2.3.4 Theoretical Framework

Most learning algorithms will perform poorly when faced with tasks having minimal data labels as well as high dimensional space. This is a familiar outcome in medical image analysis as seen before. MTL works in such instances by sharing attributes and features between tasks. Caruana (1997) best summarized the importance of MTL by highlighting it as technique that simultaneously learns tasks (parallel learning) while sharing low dimensional representations (Bib, 2019). Thus, a common assumption that is held by MTL is that tasks or their subsets, associate with each other and share information. This collaboration facilitates a joint learning process that also compares functions eventually producing optimal independent models.

This idea of jointly learning problems can be formulated as shown below. Taking $N$ as the number of supervised tasks. The training set for each and every task can be denoted as $T_n$ = ($x_i n$, $y_i n$). Where $I$ = 1: $k_n$ (kn being the number of training samples for all the tasks).

Because $x_i n$ is an element in the set of $X(n)$ i.e. $x_i n$ part of $X(n)$ and, $y_i n$ part of $Y(n)$. Then the overall problem of multitask learning can be summarized to an optimization problem as defined below:

$$\min_{w} \sum_{n=1}^{N} L(Y^{(n)}, f(X^{(n)} + \lambda \parallel f \parallel) \qquad (5)$$

$L$ is the loss function that measures the pre-task prediction error, while f is the actual multitask model. W on the other hand is the algorithm's parameter set.

## 3 DATASET & PREPROCESSING

### 3.1 Dataset

Different types of datasets were used in this study. The First dataset was Invasive Ductal Carcinoma (IDC). It was composed of 198,738 negative samples and 78,786 positive samples. IDC is one of the most prevalent of breast cancer sub-type.

The second dataset was skin cancer with a total of 3297 training samples. Out of this 1497 were benign while the remaining 1800 were malignant. The third dataset, brain MRI, was relatively smaller in comparison to the first and the second one. It was composed of 155 brain MRI images that are tumorous and 98 brain MRI images that are non-tumorous. All the datasets used in this study were obtained from various competitions listed on Kaggle.

The skin, brain tumor and IDC datasets were used in building MTL-based and TL-based models. Succinctly, the key learning task for the MTL-based model was detection of IDC and the corresponding dataset was used. The same applied to the auxiliary tasks: detection of skin and brain cancer.

Skin dataset was again used in building the TL-based model. It was used to train the base model. The learned representations from the base model were transferred and used to train a new model to detect IDC.

### 3.2 Preprocessing

All the images in this study were resized to 32x32x3. This was based on the fact that most of the tumor cells usually occupies a very small segment given a sample image. Apart from resizing the images, other image preprocessing techniques applied during training included: i) Random cropping with a padding (4), ii) Random Horizontal Flip, iii) Random Rotation and iv) Normalization with a mean and standard deviation of 0.5.

## 4 MODEL ARCHITECTURE & TRAINING

Two sets of models were built in this study; Multitask Learning (MTL) and Transfer Learning (TL) based models. These two models had a similar goal: IDC detection.

The MTL-based model was inspired by the classical research done by Caruana (1997). The model was built with IDC detection as the main task and skin and brain cancer as auxiliary tasks. TL-based model on the other hand, was built by first training a model to detect skin cancer then the learned representations were transferred to a new model which was further trained (with additional layers) to detect IDC. MTL and TL-based model's base architecture were designed based on ResNext network model.

ResNext, a deep neural network architecture, was inspired by another network model called ResNet. The ResNet network design is a deep learning model which overcame the challenges associated with training very deep neural networks. Based on its design, the output of successive layers is concatenated with the original input then fed to the next layer and the same process repeated while going deeper into the network (He et al., 2016). This design behaviour proved effective in training a neural network with over a hundred and even a thousand layers. Figure 2 shows a ResNet block.

ResNeXt leverages on a split-transform-merge strategy where branched paths are used within cells. Instead of performing convolutions on the full input feature map, the input block is projected over a series of lower dimensional representations which separately apply a few convolutional filters before merging them into the final result (Xie et al., 2017). Figure 3 captures the key difference between ResNet and ResNext Architecture.

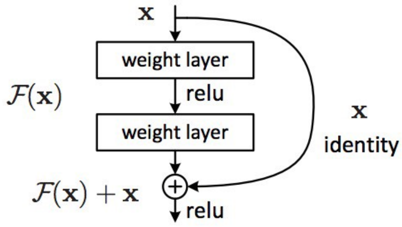


Figure 2: ResNet Block Source He et al. (2016).

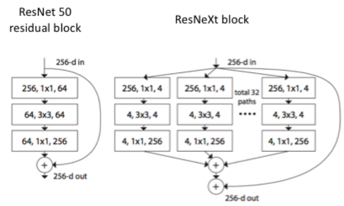


Figure 3: ResNet Vs ResNext Source Xie et al. (2017).

### 4.1 MTL-based Model

MTL-based model was designed to learn to detect IDC (key task), brain tumour and skin cancer (auxiliary tasks). The aforementioned tasks shared common base representations (parameters) with each having a unique last layer composed of a Fully Connected (FC) Layer. The MTL parameter sharing technique used in this context was hard parameter.

Figure 4 is a diagrammatic illustration of the MTL-based model. It is composed of two major parts: i) the base and ii) head. The base sub-section is composed ResNext architecture (depth: 29, cardinality: 8, base width: 64 , widening factor: 4) shared by all the tasks. The head part on the other hand is composed of three different fully connected (FC) layers: FC0, FC1 and FC2, each representing IDC, skin, brain detection model respectively. All the FC layers shares a common base while focusing on different tasks. In essence the learning signals from skin, IDC and brain affects the base weights as the head focuses on a more specific tasks. The model design was influenced by the question, *Can learning from different tasks improve model's performance?* which is theoretically and empirically underlined by Caruana (1997).

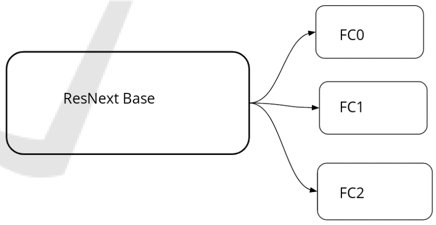


Figure 4: MTL-based Model.

### 4.2 TL-based Model

The second model which was designed in this study is TL-based model. In contrast to the first model (MTL-based), it was first trained to detect skin cancer. The learned representations were then transferred and used to train a new model-with additional layers. The new model was fine-tuned to detect IDC by adding the following layers: i) two FC's, ii) RELU , iii) Dropout and iv) Batch Norm. Figure 5 shows the high level architecture of the TL-based model. Similar to the MTL-based model architecture, the TL's base model (where knowledge is transferred from) is made up of ResNext architecture as well. The key difference is that MTL-based model is focused on multiple tasks while TL-based model simply reuses to source knowledge to enhance the new model's performance.

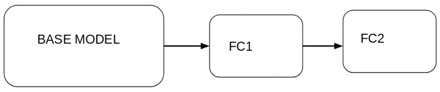


Figure 5: TL-based Model.

### 4.3 Training

The models were trained for 40 epochs. Their learning rates were varied over time (significantly lowered as the loss gradually decreases) on the range between 0.0001 - 0.01. All the experiments were carried out using PyTorch version 1.0 hosted on AWS's p2.xlarge GPU machine. Vanilla Stochastic Gradient Descent with momentum of 0.9 was used as the optimization algorithm for the two models.

## 5 RESULTS

MTL-based model was trained to detect brain tumor, skin and IDC. The model's key task was detection of IDC whereas skin and brain tumor detection were auxiliary tasks. Skin, brain and IDC sub-models shared common base but different last layers as depicted in Figure 4. Table 1 shows the result summary of the MTL-based model performance.

The TL-based model on the other was first trained on skin cancer dataset and the knowledge representations transferred (as a base model) to a new model. The latter was fine-tuned by adding two Fully Connected Layers (FC's), RELU, Dropout and Batch Norm. Table 2 summarizes the results of the TL-based model.

## 6 DISCUSSION

Multitask Learning (MTL) and Transfer Learning (TL) are key learning methods which have been widely adopted to enhance model's performance. In this study, we designed MTL-based model and TL-based model and compared their results in detection of IDC. The results of the two models were tabulated in Table 1 and Table 2 respectively.

From Table 1, the MTL-based model on IDC detection recorded cross entropy training loss of 0.26, validation loss: 0.6, validation accuracy:75.99, and training accuracy of 88.50. Comparatively, the TL-based model recorded a cross entropy training loss of 0.43, validation loss of 0.51, validation accuracy of 67.38 and training accuracy of 79.12. Moreover, the results of auxiliary tasks (skin and brain tumor) can be read from Table 1. In addition, the performance of base model used to transfer knowledge to TL-based model can also be read from Table 2.

Based on the results, the accuracy difference between MTL-based model and TL-based model on IDC detection was 8.6 on validation set and 9.37 on the training set. On comparing the loss metric of the same models, a cross entropy of 0.18 was recorded on validation set and 0.08 on the training set. Considering the accuracy of the two models' on IDC detection, MTL-based model performed was better as compared to the TL-based model.

Even though the results follow logically from the theoretical and empirical foundation presented by Caruana (1997) and Ruder (2017), more experiments need to be conducted to absolutely conclude that MTL-based models are generally better than TL-based models in the context of cancer detection. This is based on the fact that first: there are other cancer classes which were not considered in this study. Two, TL-based model has the potent to yield the quite competitive results (as shown in Table 2). Lastly, the skin dataset and brain MRI dataset used in this study was relatively small as compared to datasets which might be available for other cancer classes.

## 7 CONCLUSION

In this paper, we investigated how MTL and TL based models performance compare when applied to the IDC detection. Following the obtained results, MTL-based model recorded a better performance. TL-based model yielded relatively competitive results.

Even though the outcome of this research coincided with the theoretical and empirical underpinnings of other studies, there is still the need to con-

Table 1: MTL-based Model Results.

| **Model** | **Skin** | **Brain** | **IDC** |
|---|---|---|---|
| **Training Loss** | **0.27** | **0.6** | **0.26** |
| **Validation Loss** | **0.48** | **0.9** | **0.6** |
| **Validation Accuracy** | **75.92** | **62.5** | **75.99** |
| **Training Accuracy** | **86.90** | **60.7** | **88.50** |

Table 2: TL-based Model Results.

| **Model** | **Base Model (on skin dataset)** | **IDC** |
|---|---|---|
| **Training Loss** | **0.19** | **0.43** |
| **Validation Loss** | **0.17** | **0.51** |
| **Validation Accuracy** | **81.12** | **67.38** |
| **Training Accuracy** | **93.77** | **79.12** |

duct more experiments with different types of cancer on much larger datasets. This would allow drawing more general conclusions about the performance of MTL-based and TL-based models in the context of cancer research.

# 8 FUTURE WORK

Following the obtained results, MTL-based model yielded much better outcome compared to TL-based. In future work, we seek to research on the models' performance based other cancer sub-types and on much larger datasets. This would allow drawing conclusive remarks on TL-based or MTL-based model performance in the context of cancer research.

# REFERENCES

(2019). (1) (PDF) A brief review on multi-task learning. [Online; accessed 10. Jul. 2019].

Asgarian, A. (2019). Transfer Learning—part 1. *Medium*.

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., et al. (2019). Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: a cancer journal for clinicians*, 69(2):127–157.

Caruana, R. (1997). Multitask learning. *Machine learning*, 28(1):41–75.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Khosravan, N. and Bagci, U. (2018). Semi-supervised multi-task learning for lung cancer diagnosis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 710–713. IEEE.

Klang, E. (2018). Deep learning and medical imaging. *Journal of thoracic disease*, 10(3):1325.

Liu, X., Wang, C., Hu, Y., Zeng, Z., Bai, J., and Liao, G. (2018). Transfer learning with convolutional neural network for early gastric cancer classification on magnifiying narrow-band imaging images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1388–1392. IEEE.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI Global.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500.

Zhang, Y. and Yang, Q. (2017). An overview of multi-task learning. *National Science Review*, 5(1):30–43.