Abstract

Craniosynostosis (synostosis) is a serious disease where the sutures of a newborn’s skull fuse prematurely leading to debilitating head shape deformities. Due to the seriousness of this condition many normal infants and those with benign head shape abnormalities are referred to pediatric craniofacial plastic surgeons, leading to a high referral burden and delays in diagnosis for patients. A diagnostic delay beyond 4 months of age excludes patients from being treated with minimally invasive endoscopic procedures, leading to higher risk open surgeries. Machine learning (ML) image classifiers can enhance the triaging process of these referrals through the use of 3D images taken by a multi-camera & angle setup during patient visits. In doing so, children with synostosis can be identified earlier, qualifying them for less invasive endoscopic surgical intervention. After training a variety of convolutional neural network (CNN) models on 3D images supplemented with synthetic images using generative adversarial networks (GANs), the best-performing model was found to be a novel approach developed in our study called a multi-view collapsed 3D CNN, which achieved area under the receiver operating curves (AUROC) between 90.00-97.00% for detecting various sub-types of synostosis. These results demonstrate the ability for ML models to potentially streamline the detection of children with synostosis and help overcome challenges associated with high referral burdens for these patients.
Keywords: craniosynostosis, 3D photogtaphy, convolutional neural networks, image classification, triage, GANs

1. Introduction

Infant skull deformities have become increasingly common since the introduction of the Back to Sleep Campaign Roby et al. (2012), occurring at least once in every 60 live births. The majority of these abnormalities are related to plagiocephaly — a relatively benign condition that can be treated with simple interventions such as changing sleep positions and helmeting if severe Khanna et al. (2011). Conversely, craniosynostosis (synostosis) is a more severe condition in which one or more of the sutures in an infant’s skull fuse prematurely, causing debilitating abnormalities in skull growth and potential impacts on brain development Biggs (2003).

Distinguishing between plagiocephaly and synostosis is a technically difficult task for primary health care providers Yan et al. (2018), resulting in both conditions being referred for further consultation at high rates. The craniofacial program at our tertiary children’s hospital (Hospital for Sick Children) receives thousands of new consultations for infant skull deformities per year, with the majority being normal and plagiocephaly cases. This high volume of patient referrals results in prolonged consult visit wait-times and diagnostic delays for those with synostosis. This is of particular concern, as infants who are diagnosed with synostosis prior to 4 months of age can be treated with a relatively minimal invasive endoscopic surgery Proctor (2014). Those diagnosed after this window of time require an extremely invasive procedure in which the skull is removed from the head, reshaped, and applied back on. This open procedure is known to have a higher risk for complications such as bleeding, longer lengths of stay in hospital, and is associated with higher healthcare cost Proctor (2014). Unfortunately, approximately half of the patients assessed at our hospital will have missed this window of opportunity due to delays at both the community and hospital level. Essentially, cases of synostosis are lost amongst a sea of normal referrals. Identification of patients with synostosis early in the triage process may help expedite care for those who are at highest risk of needing surgical intervention.

Currently, our hospital triages referred patients mostly on a first come first serve basis while also considering the potential severity of the case as described by the referring provider. The variability in how synostosis patients are described in their referrals leads to an inability to consistently identify the highest risk patients. Hence, a novel standardized process for identifying high-risk patients and prioritizing them in a referral process is required.

In this paper, we hypothesize that synostosis can be distinguished from cases of plagiocephaly and those with normal head shapes using convolutional neural networks (CNNs) trained on 3D camera images of infants. In doing so, we aim to build a machine learning (ML) based synostosis triaging system to identify patients early and expedite their care towards minimally invasive surgical interventions as needed.

Although there have been other studies that have focused on this problem, we believe we are the first to use machine learning on 3D photography images to detect synostosis in infants. A similar study done by Porras et al. (2019) demonstrated very impressive results in predicting synostosis using 3D photography images, however the age population they focused on and their analyses were very different from what was done in this paper. The authors of the paper used 3D images to first extract anatomical metrics that are known to be diagnostic of synostosis and then use a machine learning model trained on those metrics (not the images) to predict if synostosis
is present. Additionally, the study included patients up to the age of 6, which makes their model less clinically relevant when solving our particular challenge of diagnosing infants prior to 4 months of age. In our paper we aimed to create an end-to-end workflow that predicts from the images directly without any manual preprocessing, and focuses only on patients under 1 year of age. Other similar studies done in this field use various image capturing modalities, such as CT scans, which generally add bottlenecks in the image capturing process and/or expose children to potentially harmful radiation Ruiz-Correa et al. (2005) Yang et al. (2011).

2. Data

2.1. Data Collection Process

Images were acquired via the 3dMDface System (www.3dmd.com). The multi-camera setup consists of 5 modular pods, each containing 3 medical grade cameras for a total of 15 cameras. These pods cover the back-left & right, front-left & right, and top-down views of a patient. The software takes the above information and creates a set of five 2D RGB images (one from each angle), as well as one 3D point-cloud. Although the 2D RGB images were of good quality, they contained a lot of background noise that made it difficult to distinguish the subject (infant head) from the rest of the image. For this reason, we have decided to only focus on the 3D dataset generated from this system. To provide a higher resolution of the shape of the head, hair was flattened by placing a single layered nylon cap on the patient’s head, keeping the ears and forehead exposed. Due to the main purpose of the project being to create an early detection system for classifying synostosis in young infants, the dataset was restricted to only contain children aged 1 year and under. Furthermore, any images of patients that had undergone surgery for synostosis have been removed from the dataset. The distribution of the remaining images is shown in Table 1. The diagnosis and corresponding label for each patient was determined after in-person clinical encounters with Pediatric Plastic Surgeons specializing in craniofacial diseases at the Hospital for Sick Children.

2.2. Labels

The primary purpose of this project was to use a machine-learning approach to predict if a subject has synostosis, plagiocephaly, or a normal head. Although this leads to 3 primary classes, synostosis itself is an umbrella term consisting of multiple sub-types; primarily sagittal, metopic, and unicoronal. These three sub-types vary greatly from each other in how they present anatomically, and each have their own thresholds for intervention. Thus, we expanded the synostosis sub-types as additional labels, resulting in 5 total classes: sagittal synostosis, metopic synostosis, unicoronal synostosis, plagiocephaly, and normal.

2.3. Data Pre-Processing

Our dataset originally consisted of 3D point-clouds. Although there have been great breakthroughs on training classification models using 3D point-clouds, such as PointNet Qi et al. (2017), there are still very few cases where
such a modality is applicable. Point-clouds still face the constraint of requiring a uniform input shape for model training. However in our dataset, the shape of the raw point-cloud images varied greatly, and we did not feel comfortable subsampling and removing any part of the dataset randomly to achieve a uniform shape (which is a common technique when dealing with point-cloud based models). For this primary reason, we decided to move away from point clouds and focus on a voxel-based approach.

The point-cloud images were converted to 3D voxels using binvox Min (2004 - 2019). The resulting dataset consisted of 3D images of the shape (128 x 128 x 128 x 1), where each point represents whether a voxel exists or not. The images were also mirrored along the vertical axis to help naturally augment the dataset. It is of importance to note that the validation and test datasets did not undergo any augmentations, and no flipped images exist within either datasets.

2.4. Dataset Split

The dataset was split into training/validation/test sections, with a 60%/20%/20% split of patients. This was done to prevent any overlap of images from the same patient between the training section and the validation/test sections. To account for the overall class imbalance in the dataset, each class in the training dataset was upsampled to create an even distribution, while the validation and test datasets were left untouched.

3. Methods

All models were developed under Python v2.7.12, running Tensorflow Abadi et al. (2015) v1.14.0. All models were optimized using ADAM Kingma and Ba (2014), and ran for 350 epochs. Dropout and EarlyStopping were utilized to address model overfitting.

3.1. 3D CNN Model

A standard 3D ResNet18 (https://github.com/JihongJu/keras-resnet3d) model architecture was trained using 3D voxel data. This model took as input a single 3D voxel image, with the model output consisting of one of five predictions for sagittal synostosis, metopic synostosis, unicoronal synostosis, plagiocephaly, or other/normal. One strong benefit of using the 3D data was the depth information it contained along with the spatial relationships between key anatomical landmarks. We believe this will help the model not only understand the different types of deformities, but also learn to gauge its severity.

3.2. Collapsed 3D CNN Model

There were concerns regarding how limited the dataset was with respect to the wide 3D problem space. In an effort to reduce the problem space to a 2D space, while still maintaining some aspects of the original depth and spatial information, a novel method was developed (henceforth referred to as the collapsed 3D CNN model). This method takes the existing 3D voxel of the head and cuts it along the transverse view, and only keeps the top half of the head. This subsection of skull contains the most informative anatomical landmarks for diagnosing synostosis. The slices from this new view were then subsequently aggregated (summed) together to create a single 2D image, a top-down view from the collapsed transverse section. In this 2D space, the colour (i.e. the brightness) is a direct indication of how dense that portion of the head is. This transformation drastically reduces the problem space (from 3D to 2D), while still holding some latent representation of the relative 3D depth across key anatomical aspects of the skull (Figure 1a). A breakdown of this new view for each class can also be found in Figure 1b.
The final 2D image created from this pre-processing approach is very different from the original 2D images generated by the system. First, the pixel values from the collapsed generated image reflect a latent representation of the depth of the skull at that point, which gives us a better understanding of which areas of the skull contain deformities, even in this 2D space. The original RGB images created by the system have no relation to depth, which makes them much harder for a model to train on. Additionally, the large amount of background noise and variation in the original RGB images introduces noise that are very difficult to isolate from, which would again cause the model to not converge efficiently.

The collapsed 3D model consists of a very shallow network, with 2 CNN layers and 2 subsequent fully-connected layers. This configuration was carefully selected due to the limited amount of training images that were available. Similar to the 3D CNN model, the collapsed 3D model was trained for 350 epochs, with EarlyStopping used to prevent overfitting.

3.3. Collapsed 3D Image Generation

Unlike most medical projects, our project suffers from having too few control/normal images, as opposed to too many. This was found to cause a significant drop in performance for predicting our other/normal class. A reasonable solution to this can be to reduce the number of images in the rest of our classes to reach a more uniform distribution; however, we found that to significantly impact the performance of our model overall.

Instead, we opted to augment our normal class in order to increase the number of images. To do so, we trained a standalone WGAN GP model Gulrajani et al. (2017) on the normal class training set images, and used that to generate new 'normal' images. This model was trained for 15000 epochs, after which it was tasked to generate 100 images that were classified by the discriminator as a 'real' image
(a) A real image from the Collapsed 3D processing

(b) Five images generated by the WGAN GP model

Figure 2: Real vs WGAN GP generated images

(i.e. were able to fool the discriminator). The generated images can be seen in Figure 2.

3.4. Multi-view Collapsed 3D Image Classification

A logical limitation with our current design is that we are only viewing the new 2D image from a single view, the top-down view. In an effort to widen our field of scope, we decided to additionally slice the 3D image across the coronal view, which gives us a back view of the skull. This helps provide us with two different 2D images from the original 3D image, providing us with more viewpoints of the skull while keeping the problem space constrained to 2D. The CNN model was then altered to accept two 2D images as input, and provide a prediction for the same 5 classes mentioned before.

4. Results

4.1. 3D Image Classification

Our 3D CNN model leverages information pertaining to depth, contouring, and space between key anatomical relationships, leading to AUROCs for predicting sub-types of synostosis between 93-97%. These excellent results are overshadowed by the model’s high false negative rates as seen in Figure 3a, leading to cases of synostosis being under-recognized and inadvertently bumped backwards in priority.

4.2. Collapsed 3D Image Classification

The collapsed 3D CNN model saw improved performance compared to our traditional 3D CNN model, due to it containing such dense information while being restricted to a 2D data space. AUROCs for predicting sub-types of synostosis ranged from 89-99% with true positive rates ranging between 79-91%. When identifying cases of plagiocephaly and other/normal patients AUROCs were 87% and 74%, respectively (as can be seen in Figure 3b). This model out-performs our 3D CNN model, as it provides with much more confident synostosis predictions while minimizing both false-negative and false-positive rates.
4.3. Collapsed 3D Image Generation

Our WGAN GP model was trained for 15000 epochs, after which it was tasked to generate 100 images that were classified by the discriminator as a ‘real’ image. These 100 new generated ‘normal’ images were then appended to our training dataset, and the CNN model was re-trained. Results for this can be found under Table 2 and Figure 3c. The generated images helped not only improve the true positive rate of our normal class by 20%, but also increased the AUROC scores between 1%-4% across all classes. Although adding in the generated images did result in improvement to the overall model performance, there was an increase in false negatives for the sagittal subclass, although was still much lower relative to the original 3D model. This model did however obtain a lower false negative rate for the metopic sub-class.

4.4. Multi-view Collapsed 3D Image Classification

Our multi-view approach achieved meaningful improvement over all previous models. Figure 3d shows the confusion matrix for this new analysis, showing true positive rates increased by as much as 13% compared to the single-view with generated images model. Additionally, AUROC scores have also shown some increase, with the synostosis sub-classes hitting 90-97%, plagiocephalhy at 89%, and other/normal at 81%.

4.5. Next Steps

More research will be dedicated into modifying our WGAN GP model to be able to generate multi-view 2D images for all classes, which we believe will boost performance further enabling our ultimate goal of clinical deployment.

5. Discussion

Synostosis is a relatively rare but serious time sensitive problem that is increasing in prevalence, occurring 6.4 in every 10,000 live births Cornelissen et al. (2016). To any individual community clinician the low frequency of exposure to these cases creates a diagnostic challenge when differentiating synostosis from the large number of infants born with benign head shape deformities, as both patient populations present in a similar fashion. This contributes to delays in referrals for patients with synostosis who are initially thought to have plagiocephaly along with issues in over-referral of normal and benign head shape deformities. In any given state or province in North America, there typically is one major pediatric hospital with the expertise to formally diagnose and treat patients with synostosis. As a result, majority of referrals related to head shape concerns are directed towards these limited centers. This further contributes to significant referral burden and an inability to consistently provide timely care. These clinical and logistical challenges cause many patients to not qualify for lower risk, lower cost, minimally invasive procedures which can only be performed on patients 4 months of age and younger. Hence, a solution is required to more rapidly identify cases of synostosis at both a community and territory care level.

For those patients who are referred, the main bottleneck in the assessment pathway is the availability of an expert craniofacial plastic surgeon to confirm whether the disease is present. The positive results obtained from our collapsed 3D CNN models demonstrate the potential for ML to assist in detecting and triaging infants with synostosis, enabling non-invasive diagnostics to begin earlier while waiting for the surgeon appointment date. In doing so, a risk score can be applied to all referred patients allowing for the prioritization
Table 2: Outcome metrics for the 3D & Collapsed CNN models

<table>
<thead>
<tr>
<th>Model</th>
<th>Class</th>
<th>Sub-type</th>
<th>Precision</th>
<th>Recall</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Synostosis</td>
<td>Sagittal</td>
<td>76.0%</td>
<td>73.1%</td>
<td>97.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metopic</td>
<td>80.8%</td>
<td>80.8%</td>
<td>93.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unicoronal</td>
<td>62.9%</td>
<td>84.6%</td>
<td>93.0%</td>
</tr>
<tr>
<td></td>
<td>Plagiocephaly</td>
<td>-</td>
<td>83.0%</td>
<td>34.6%</td>
<td>83.0%</td>
</tr>
<tr>
<td></td>
<td>Other/Normal</td>
<td>-</td>
<td>40.6%</td>
<td>50.0%</td>
<td>78.0%</td>
</tr>
<tr>
<td>Collapsed 3D</td>
<td>Synostosis</td>
<td>Sagittal</td>
<td>88.6%</td>
<td>91.2%</td>
<td>97.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metopic</td>
<td>68.8%</td>
<td>81.5%</td>
<td>93.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unicoronal</td>
<td>84.6%</td>
<td>78.6%</td>
<td>89.9%</td>
</tr>
<tr>
<td></td>
<td>Plagiocephaly</td>
<td>-</td>
<td>81.6%</td>
<td>73.8%</td>
<td>87.0%</td>
</tr>
<tr>
<td></td>
<td>Other/Normal</td>
<td>-</td>
<td>47.4%</td>
<td>45.0%</td>
<td>74.4%</td>
</tr>
<tr>
<td>Collapsed 3D w/ GAN images</td>
<td>Synostosis</td>
<td>Sagittal</td>
<td>85.7%</td>
<td>88.2%</td>
<td>98.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metopic</td>
<td>72.4%</td>
<td>77.8%</td>
<td>94.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unicoronal</td>
<td>91.7%</td>
<td>78.6%</td>
<td>93.0%</td>
</tr>
<tr>
<td></td>
<td>Plagiocephaly</td>
<td>-</td>
<td>84.2%</td>
<td>76.2%</td>
<td>88.1%</td>
</tr>
<tr>
<td></td>
<td>Other/Normal</td>
<td>-</td>
<td>56.5%</td>
<td>65.0%</td>
<td>78.1%</td>
</tr>
<tr>
<td>Multi-view Collapsed 3D</td>
<td>Synostosis</td>
<td>Sagittal</td>
<td>83.3%</td>
<td>96.2%</td>
<td>96.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metopic</td>
<td>84.6%</td>
<td>84.6%</td>
<td>92.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unicoronal</td>
<td>88.8%</td>
<td>92.3%</td>
<td>96.0%</td>
</tr>
<tr>
<td></td>
<td>Plagiocephaly</td>
<td>-</td>
<td>84.2%</td>
<td>61.5%</td>
<td>92.0%</td>
</tr>
<tr>
<td></td>
<td>Other/Normal</td>
<td>-</td>
<td>60.7%</td>
<td>65.4%</td>
<td>80.0%</td>
</tr>
</tbody>
</table>
of patients with the highest risk of needing surgery.

Our best performing model was our multi-view collapsed 3D CNN model which obtained promising AUROC scores (90-97%) for diagnosing sub-types of synostosis along with associated high true positive rates (85-96%) and relatively low false negative rates (Figure 3d). We suspect this model obtained the best performance for both clinical and technical reasons. The assessment of head shape deformities requires clinical assessment of anatomical skull landmarks from multiple views during a typical plastic surgery clinic encounter. Depth, contouring, and the geometric space between anatomical landmarks are richly captured within our 3D dataset with much of this information being preserved in the multi-view...
collapsed 3D model data. This allowed for the addition of more clinically relevant images into our model while also benefiting from the reduced feature input space when compared to traditional 3D image inputs.

Our 3D CNN model performed well overall, however, it was plagued with relatively low AUROCs and high false negative rates limiting its ability for clinical implementation. This is an important consideration as false negative cases risk patients with synostosis being bumped back in the triage queue. This will further add to diagnostic delays and is therefore critical that the final implemented model have a low false negative rate. The low ceiling hit when training this model was likely due to the high dimensionality of our 3D data and the relatively small number of patients within our dataset. To overcome this, an attempt was made to reduce the dimensions of our 3D dataset while preserving some of the enhanced anatomical relationship data contained within it. This enabled us to build a superior collapsed 3D CNN model that outperformed our traditional 3D CNN and appears to be a useful approach when working with small 3D datasets. In a data rich environment it is not clear if our collapsed 3D model will continue to outperform traditional 3D CNN approaches. Clarifying this will be a focus of future work moving forward.

For applications where multiple views of the skull may not be readily available and training data limited as is often the case when working in niche medical domains, our single view collapsed 3D CNN model supplemented with synthetic GAN images for training appears to be an excellent alternative approach. Such applications may include scaling to mobile devices where 3D capture of a single view may create a more user-friendly experience when compared to trying to capture multiple angles of a child’s head.

Given the relatively limited size of our dataset, we cannot confidently state that our models will generalize broadly. Rather this work is important for demonstrating the key role ML can have in augmenting the triage and diagnostic pathways for patients with synostosis, and highlights the need for future collaboration and data sharing amongst pediatric centers to advance this work forward. Completing in-situ prospective clinical evaluation both at our local pediatric hospital and community partner sites will be an essential next step for ensuring robust model validation is completed prior to integration into clinical workflows. In this process, a quantitative assessment of patient-centered outcome metrics such as time to surgery, incidence of surgical complications, length of stay in hospital, and patient satisfaction can be assessed along with overall healthcare costs.

An additional next step in this research will focus on adapting models to predict on 3D image data obtained from mobile devices. This will enhance our ability to translate to community providers. With the rapid advancement of 3D camera technology in commercially available mobile devices, there is an opportunity to focus translation efforts to community clinicians. In doing so, risk stratification of patients with head shape deformities can be completed earlier in the referral pathway allowing for more children to receive less invasive endoscopic surgery for synostosis.

6. Conclusion

3D images can be used to train CNN models in order to develop an effective synostosis triaging system. These ML based models may potentially allow for the translation of diagnostic knowledge from pediatric centres to community care providers, enabling early identification of patients with synostosis. In doing so, we can likely improve the quality of care provided to infants with synostosis significantly, while simultaneously reducing associated healthcare costs.
References


Han Yan, Taylor J Abel, Naif M Alotaibi, Melanie Anderson, Toba N Niazi, Alexander G Weil, Aria Fallah, John H Phillips,