2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Bottle-Feeding Intervention Detection in the NICU

Yasmina Souley Dosso, *Student Member, IEEE,* Kim Greenwood, JoAnn Harrold, and James R. Green, *Senior Member, IEEE*

Abstract— **Video-based monitoring of patients in the neonatal intensive care unit (NICU) has great potential for improving patient care. Video-based detection of clinical events, such as bottle feeding, would represent a step towards semi-automated charting of clinical events. Recording such events contemporaneously would address the limitations associated with retrospective charting. Such a system could also support oral feeding assessment tools, as the patient's feeding skills and nutrition are pivotal in monitoring their growth. We therefore leverage transfer learning using a pretrained VGG-16 model to classify images obtained during an intervention, to determine if a bottle-feeding event is occurring. Additionally, we explore a data expansion technique by extracting similar-feature images from publicly available sources to supplement our dataset of real NICU patients to address data scarcity. This work also visualizes and quantifies the gap between the source domain (ImageNet data subset) and target domain (NICU dataset), thereby illustrating the impact of expanding our training set for knowledge transfer proficiency. Results show an increase of over 18% in sensitivity after data expansion. Analysis of network activation maps indicates that data expansion is able to reduce the distance between the source and target domains.**

I. Introduction

Continuous monitoring of patients in the neonatal intensive care unit (NICU) has been increasingly studied in recent years. Many research groups have implemented non-contact monitoring systems for vital signs monitoring [1], [2], motion detection [3], [4], and pain assessment [5], to name a few. More recent non-contact studies have further analyzed the NICU environment by tracking the face of the patient [6], detecting patient presence in the bed [1], [7] and detecting clinical interventions [1]. During continuous monitoring, clinical interventions can sometimes pose a problem in the development of non-contact monitoring systems since clinical staff can naturally occlude portions of the patient. In many cases, intervention periods are actually excluded from analysis [1]–[4]. Clinical interventions may represent pivotal moments in newborns' continuous care in the NICU; further investigation of these events is therefore warranted. Routine care events include diaper change, feeding, checking temperature, checking vital signs, weighing, changing sensors, to name a few. All interventions must be carefully documented in the patient's chart indicating the date, time, personnel in charge, and any relevant details [8]. Such documentation is important since it is the primary source of communication between clinical staff. Ideally, all events would be documented contemporaneously. Realistically, the elevated nurse workload, combined with patient acuity requiring more time for patient care, often results in retrospective charting. Retrospective documentation is problematic since it is often incomplete or inaccurate. Therefore, a clinical assistive tool to automatically identify and chart interventions could help address this issue.

Previous studies have established the importance of monitoring the oral feeding process (breast-feeding and bottle-feeding) in newborns less than 6-7 months old [9]–[11]. In fact, newborns' acquired feeding skills are crucial in these first few months as they directly impact their nutrition and brain development, especially for those born prematurely or those in critical health condition. Assessment tools have then been implemented to help guide nurses or parents in evaluating the newborn's feeding skills [12]–[14]. This includes evaluating the newborn's state when bottle-fed, such as muscle tone, readiness, sucking, swallowing, breathing, fatigue, and oral-motor patterns. These assessment tools can help in deciding when the patient can safely be discharged from the hospital, and for transitioning to solid food in the NICU or at home [11].

This study focuses on video-based detection of bottle-feeding interventions as a step towards automated clinical documentation in the NICU, while supporting neonatal oral feeding assessments. To this end, we aim to identify in each image whether a bottle-feeding event is occurring or not. Recent neonatal monitoring studies have reported detection of clinical interventions; however, no specific detection of a particular intervention types, such as feeding, has been investigated [1].

We herein leverage transfer learning on a pretrained neural network, VGG-16 [15], to classify "bottle-feeding" vs "no-feeding" events in an intervention image. In its most basic definition, *transfer learning* describes a process in which some pertinent information is passed from a certain source domain to a target domain [16]. In cases where both domains have labeled data, this procedure called *inductive transfer learning* is often performed by transferring pre-learnt weights from a model to another domain [17]. Both domains are assumed to be similar, so that key features can be reused from a source task to perform a new target task, instead of learning all model weights from scratch. Model parameters can then be fine-tuned to be tailored for the target task. Often the volume of available data in the target domain is significantly smaller than

This study was funded by the IBM Centre for Advanced Studies and by the grant from the Natural Sciences and Engineering Research Council [CRDPJ 543940-19].

Y. Souley Dosso and J.R. Green (corresponding author) are with the Department of Systems and Computer Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6F, Canada (email: yasmina.souleydosso@carleton.ca, jrgreen@sce.carleton.ca).

K. Greenwood Director of Clinical Engineering, Children's Hospital of Eastern Ontario, and an adjunct professor in the Department of Mechanical Engineering, Faculty of Engineering, University of Ottawa, 75 Laurier Ave. E, Ottawa, ON, K1N 6N5, Canada (e-mail: kgreenwood@cheo.on.ca)

J. Harrold is with Neonatology, Children's Hospital of Eastern Ontario, 401 Smyth Rd, Ottawa, ON, K1H 8L1, Canada. (e-mail: jharrold@cheo.on.ca.)

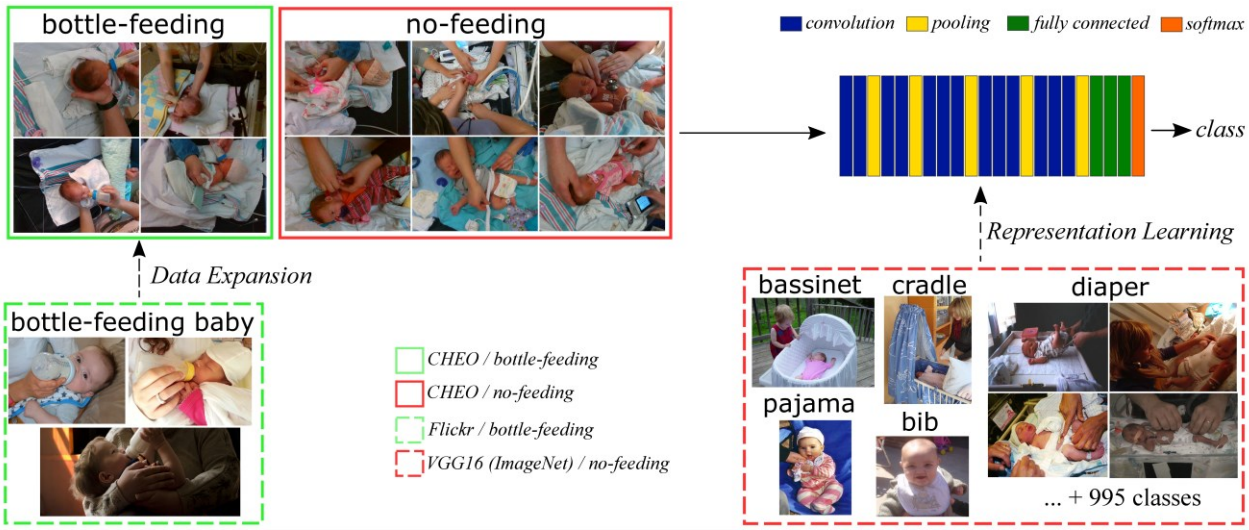


Fig.1 Transfer learning model trained on VGG-16 using a subset of ImageNet [15], [21] with data expansion from publicly available images [24]*. *Due to licensing issues, the figures portrayed were obtained from https://pixabay.com/ and only shown here for illustration purposes.

the source domain. Transfer learning can then help in cases of data scarcity, which often occurs in medical applications due to time, cost, and privacy constraints. Recent advances in convolutional neural network (CNN) models have led to significant improvements in image analysis tasks, such as image classification [18], object detection [19], and scene recognition [20]. These advances have been attributed to improved algorithms, increased computational power (especially GPU), and increased access to labelled image datasets. Multiple studies have leveraged the ImageNet dataset comprising more than 14 million annotated images, where a subset of ~1 million images across 1000 different classes is often used to train networks [21]. Given the large amount of data, this dataset has been widely used for implementing deep image classification models, resulting in powerful state-of-the-art deep neural networks including the VGG-16 model. Examples of classes used to train this network include common objects such as *bicycle*, *car*, *dog*, and *table*.

From the ImageNet dataset used to train the VGG-16 model, some classes included the cooccurrence of a baby and an adult, such as *cradle*, *diaper*, or *bib*. These object classes are present in both our positive and negative classes, however, the additional occurrence of a nursing bottle would classify our image as "bottle-feeding" if the nursing bottle is present, and "no-feeding" otherwise. Can a network, pretrained in a source domain devoid of bottle-feeding events, efficiently classify bottle-feeding events? Transfer learning can partially address the domain gap [17], but we have very limited data available in the target domain. We herein address this question by adding a third domain comprising images extracted from publicly available sources, similar in key features to a bottle-feeding intervention image. This supplemented data domain will help bridge the gap between the source and target domains for knowledge transfer proficiency, as depicted in Fig. 2. To this end, we investigate how the knowledge acquired from millions of images in a source domain, complimented by an expanded data domain, can be transferred to a significantly smaller dataset in the target domain. We evaluate the impact of data expansion in transfer learning to address data scarcity in the target domain. We additionally visualize and quantify that gap to demonstrate the influence of the data expansion domain.

II. Methods

A. Transfer Learning & Data Expansion

As previously mentioned, to perform classification of "bottle-feeding" vs "no-feeding" events, we selected the pretrained VGG-16 model due to its strong performance on previous image classification tasks [18]. The model comprises 13 convolutional layers, often referred to as the "feature extraction" layers, followed by three densely connected layers responsible for arriving at a final classification. Model parameters were tuned using preliminary experiments. Training of images involved a mini-batch size of 32 with a learning rate of 1e-5 over 20 epochs. Due to a class imbalance among images, a weighted classification layer was used to emphasize the minority class. The model is evaluated using 5-fold cross-validation where different patients were selected per fold. Additionally, given that "bottle-feeding" event were observed in only 6 out of 27 patients, these patients' data were distributed separately among the five distinct folds (one fold included two of these patient data).

During model training, data augmentation is used to improve model performance and generalization [22], [23]. Common augmentation techniques aim to create synthetic copies of the original images through image transformation. These include translation, scaling, reflections, rotation, or shearing. Training images were therefore augmented using reflections along the X and Y axis, and rotations from 0-360 degrees. Due to the nature of our dataset, where objects of interest are often small or can be found near the edges of the image, we refrained from performing translation, scaling or shearing transformations.

While traditional data augmentation produces synthetic copies of original images, we also explore a data expansion approach to extract similar-feature images from external sources to further supplement the training dataset. As previously mentioned, collecting and labelling clinical data is a laborious task due to ethics protocols, low patient recruitment rates, equipment cost, extensive data collection

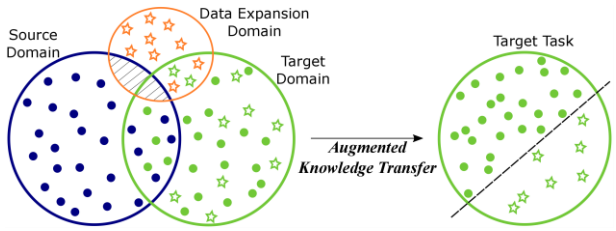


Fig.2 Similar-Feature Domain Expansion. Dots represent the negative class "no-feeding" and stars the positive class "bottle-feeding". Data expansion domain is proposed here for augmented knowledge transfer.

procedures followed by data storage, management, security, and privacy, to name a few. This often results in unique and rich data, albeit in limited quantity. We address this issue by gathering publicly available images from Flickr [24] by carefully searching and curating images showing similar objects and contexts. To this end, we obtain supplemented data from similar-feature images sufficient for performing binary classification in the target domain when the source domain is significantly deficient in data from one of these classes.

The BFID model (bottle-feeding intervention detection) was trained and tested on our original dataset. Then, the BFID*exp* (BFID with data expansion) model was trained on our expanded dataset, including Flickr images, and tested on our original dataset. Model performance is compared before and after data expansion. As a baseline method, we tested our dataset on the pretrained VGG-16 model using baby- and bottle-related classes drawn from the model's 1000 classes. These classes were extracted from a more complete list of words in ImageNet, which was structured according to the WordNet hierarchy [25]. WordNet is a lexical database of English words grouped into synsets or synonym sets if they share similar concept and semantic relations. Pictorially, these relations can therefore be demonstrated in a tree map, and a subset of this tree highlighting a few classes of interest are depicted in Fig. 3. Here, we have added a *feeding bottle* class to show word similarity. A conceptual relation can be drawn from the tree map, and we additionally visually inspected a subset of VGG-16 images from related classes to inform on feature-based relations. These classes were selected due to the environment, a piece of clothing, or an object typically seen with babies. To this end, we can curate a list of VGG-16 classes related to our data, as demonstrated in Fig. 3. From both representations, we can clearly see the close conceptual-semantic relationship between *crib*, *cradle*, and *bassinet* class, while *diaper*, *pajama*, and *bonnet* share a feature relationship to baby-related images. Most of these corresponding VGG-16 images contained a baby and sometimes an adult present but no nursing bottle, thereby similar to our "no-feeding" class. As for the "bottle-feeding" class, bottle-related classes such as *water bottle*, *pop bottle*, *beer bottle*, and *wine bottle* shared semantic relationships with each other and close relations to a feeding bottle object. However, since very few VGG-16 images contained our baby + reaching hand + bottle condition, this pretrained model contained negligible association with our bottle-feeding images. As discussed below, the baseline prediction model labels an image as "bottle-feeding" if any of the bottle-related object classes are detected in the image.

All models are evaluated using the following performance metrics among the total number of images per fold, *n*, where the positive class corresponds to "bottle-feeding" events;

$$Sensitivity = TP/(TP+FN) \quad (1)$$

$$Specificity = TN/(TN+FP) \quad (2)$$

$$Precision = TP/(TP+FP) \quad (3)$$

$$Accuracy = (TP+TN)/\sum n \quad (4)$$

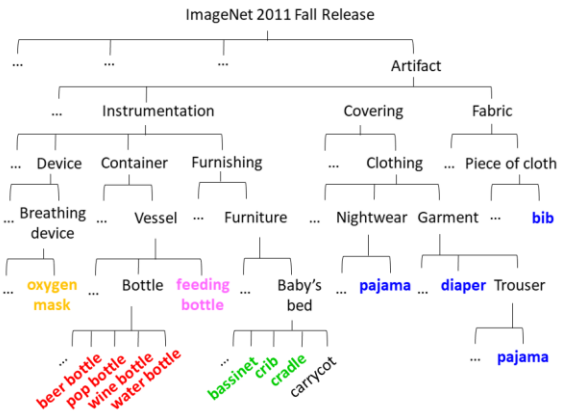
$$F1\ score = 2\left(\frac{precison \times sensitivity}{precision+sensitivity}\right) \quad (5)$$

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

Models were trained on MATLAB using Compute Canada resources, running on the *Cedar* v100l cluster [26].

B. Domain Distance Mapping

As previously mentioned, transfer learning can be useful in cases where the source domain contains a large amount of labeled data, to extract and transfer that knowledge to a smaller amount of labeled data. To this end, we repurpose the source task (classification of 1000 classes from VGG-16 model) to a new target task (classification of bottle-feeding interventions). Both domains are typically assumed to be similar; however, it is not always the case. This study presents such a scenario where the dataset used to train the VGG-16 model shares more similarity with the "no-feeding" class than the "bottle-feeding" class, as qualitatively demonstrated by Fig. 1. Although both classes share similar features from the baby, nurse, and overall bed environment, the principal distinction remains in the presence or absence of a nursing bottle. Supplemental training data using similar-feature images including a nursing bottle



Related VGG-16 Classes			Visual Description
Baby-Related	*Environment*	bassinet	Empty object or baby in the bassinet. Person (with visible face) standing by the bassinet.
		crib	Empty object or baby in the crib.
		cradle	Empty object or baby in the cradle. Person (with visible face) standing by the cradle.
	Clothing	diaper	Baby wearing a diaper or image of a diaper object. Person (face visible or not) changing a baby's diaper.
		pajama	People of all ages wearing pajamas or picture of pajamas.
		bib	Baby wearing a bib often held by an adult or empty object.
	Object	oxygen mask	Empty object or adult wearing oxygen mask. Some babies on ventilation support.
Bottle-Related	*Object*	pop bottle	Empty object or collection of objects on tables or in store.
		water bottle	Empty object or collection of objects. Some adults drinking from a water bottle.
		wine bottle	Empty object or collection of objects in wine cellar or rack.
		beer bottle	Empty object or collection of objects on tables/shelves. One picture of baby and toddler playing with beer bottles.

Fig.3 Concept relations from tree map of some VGG-16 classes. Feature relations from description of 7 baby- and 4 bottle-related classes.

can then bring both distributions closer together. The distance between domains can in fact be estimated for the BFID and BFID*exp* models. Measuring distances between domains is commonly performed in domain adaption, which is an unsupervised approach to transfer learning used when the source domain has labeled data, but the target domain does not [17], [27]. Domain adaptation seeks to minimize the gap between domain distributions during training by learning shared key features. More recently, this technique has also been used in multi-source domain adaptation where labeled data originate from multiple sources [28]. We here leverage this concept to measure the distances from the source domain to the domains of BFID or BFID*exp* models. Doing so demonstrates how the similar-feature data expansion can help narrow the distances between the source and target domain.

To quantify and visualize the domain distances, feature maps from the BFID and BFID*exp* models are individually extracted. These correspond to the activation obtained across all training samples and maps these data to the feature space. In order words, the stronger an activation in an area of the image, the greater the number of detected features in that area. Given that activation maps are provided as a greyscale image, the Otsu thresholding technique is used to differentiate between high and low intensities. The resulting blobs represent areas of heightened activation. Since the principal difference between the "bottle-feeding" and "no-feeding" events is the presence of a nursing bottle, we can estimate the domain distance within the "bottle-feeding" class as the distance between the centroid of the nearest detected blob in the feature map and the actual nursing bottle object. The closer the blob to the bottle, the closer the domains.

To measure domain distances, all nursing bottles in our image dataset were first manually segmented and represented by their centroid. The domain distance is measured by the Euclidean distance (7), (8) between the bottle and feature map centroids, since previous domain adaptation studies showed negligible difference in domain distance mapping when using different distance metrics [27], [28]. When more than one activation is detected (*i.e.,* could detect two different objects) the closest activation to the object is selected.

Domain distance mapping can also be visualized per image by overlaying feature maps, gold standard bottle centroid, and distances on the original image. By visualizing these data, we can examine the impact of expanding our model using similar-feature data by evaluating changes in activations from feature maps. Domain distance mapping is evaluated using the *%closer* performance metrics is the percentage of closer bottle-activation distances in BFID*exp* compared to BFID among bottle-containing images. Additionally, the *pixelDistance* (9) calculates how much closer the activation is to the bottle object in number of pixels using the following metrics:

$$dist_{BFID} = \sqrt{(centroid_{BFID} - centroid_{bottle})^2} \quad (7)$$

$$dist_{BFIDexp} = \sqrt{(centroid_{BFIDexp} - centroid_{bottle})^2} \quad (8)$$

$$pixelDistance = dist_{BFID} - dist_{BFIDexp} \quad (9)$$

C. Dataset

Data from 27 patients admitted to the NICU of the Children's Hospital of Eastern Ontario (CHEO) were used in our dataset. This study was approved by the Research Ethics Boards of both the hospital and Carleton University. As part of a larger overarching study, six hours of data were collected per patient while a researcher carefully annotated all events occurring at the patient's bedside using a custom Android App [29]. The start and stop times for all intervention events were annotated, including bottle-feeding interventions. A depth-sensing camera, the Intel RealSense SR300 camera [30] was securely mounted at the top of the bed (incubator, crib, or overhead warmer) to record the patient and the NICU bed environment. To ensure variations among images, one image was extracted every 30 seconds. All image sizes are 480x640 and only the color data from the camera were analyzed. Given that our dataset shares similar features (e.g., patient present, a hand from the nurse or parent reaching into the frame, NICU bed environment), an image is classified as "bottle-feeding" if a nursing bottle is present at or near the patient's mouth. If the nursing bottle is absent, the image is classified as "no-feeding". Bottle-feeding events were only seen in six of out of the 27 patients. "No-feeding" events from these six patients were also extracted during other interventions. Other patients were fed by nasogastric tube or breast-fed. The complete dataset is summarized in Table I. To supplement these hospital-based images, we extracted Flickr images showing similar objects and context in the scene using the search word "bottle feeding baby", as depicted in Fig. 1. To simulate our CHEO data, a bottle-feeding image was included if it contained a baby, a hand reaching into the frame, and a nursing bottle at or near the baby's mouth. The image environment could differ, where the baby would be placed on a pillow, blankets, a cradle, a feeding table, a baby bouncer, or in someone's arms. Images showing an adult person's face were excluded to closely simulate the NICU bed environment.

TABLE I. DATASET BREAKDOWN

Class	**Data source**			
	CHEO		*Flickr*	
	#images	*#patients*	*#images*	*#subjects*
Bottle-feeding	73	6	60	60
No-feeding	1187	27	0	0
Total	**1260**	**27**	**60**	**60**

TABLE II. BOTTLE-FEEDING INTERVENTION DETECTION

Model	**Evaluation Metrics (%)**					
	Sens	*Spec*	*Prec*	*Acc*	*F1*	*MCC*
Base	09.59	**98.90**	35.00	**93.73**	15.05	15.88
BFID	32.88 ±1.94	91.41 ±0.72	**19.10** ±1.28	88.02 ±0.64	**24.14** ±1.28	18.93 ±1.44
BFID *exp*	**51.51** ±4.06	85.96 ±3.33	**18.94** ±3.68	83.96 ±3.13	**27.54** ±4.10	**24.11** ±4.44

III. RESULTS

In this section, transfer learning results for all models are reported. In particular, the impact of data expansion on the knowledge transfer using the pretrained VGG-16 network is demonstrated. The domain distance mapping concept is finally presented to support and further explain our findings.

A. Transfer Learning & Data Expansion

As a baseline, the VGG-16 model was directly applied to our dataset and the top five predicted object classes were extracted since multiple objects can be found in the scene. This 1000-classification model outputs seven baby-related object classes and four bottle-related object classes. If the predicted object classes contained baby-related AND bottle-related classes, they were classified as "bottle-feeding". Otherwise, they were classified as "no-feeding".

In comparison with the two transfer learning models, the baseline method performs quite poorly, as depicted in Fig. 4. Unsurprisingly, the baseline model has high specificity and accuracy values, strongly suggesting that the model is classifying images as the "no-feeding" class and cannot detect bottle-feeding events. Although no bibs were used in clinical settings, that concept was useful due to association with babies, but it only appeared 1.9% of the time among the top-5 predictions. Similarly, the *oxygen mask* class included a person wearing a breathing device and some images of babies on ventilator support leading to this class being predicted for 40.6% of images. The most frequently detected baby-related class was *diaper* (77.4%), while *water bottle* was the most frequent bottle-related class (1.6%).

In comparing with BFID and BFID*exp* models, results demonstrated a significant increase in performance after transfer learning, and even further improvement after similar-feature data expansion is applied. As displayed in Table II transfer learning results overperform the baseline and overall results are better for the BFID*exp* compared to BFID, especially in sensitivity (18.63% increase) and F1-score (3.4% increase). These two metrics are most pertinent in evaluating our methods, given the high class imbalance and the greatest concern in detecting bottle-feeding events.

These findings thus corroborate how the shortage of "bottle-feeding" images used in training the VGG-16 model impacts the knowledge transfer ability to our classification task. Given that the "bottle-feeding" and "no-feeding" classes share many similar features, distinguishing between the absence or presence of the nursing bottle object is a difficult task to achieve. We have however demonstrated that our similar-feature data expansion technique can solve this issue.

B. Domain Distance Mapping

To evaluate the distance between the source and target

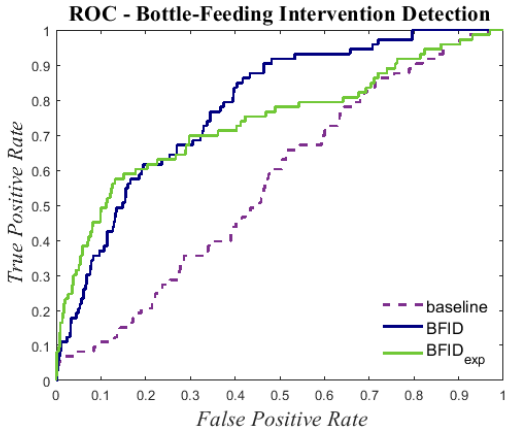


Fig.4 Bottle-Feeding Intervention Detection Results.

domain, we opted for a heuristic approach where we measure the distance between the object of interest (nursing bottle) and the models' strongest activation from the feature map. When identifying key features of an object, we can naively consider it as a whole or focus on the more salient parts. For example, when seeing a torch, we typically focus our attention on the flame, not the handle. Similarly, we explore if the attention in a nursing bottle is focused on the whole bottle or the most salient part, i.e., the bottle cap. Both objects are manually annotated for evaluation.

Results reveal that the BFID and BFID*exp* models focused more on the bottle cap than the entire bottle. This suggests that the bottle cap shows greater saliency information attributed to the bottle object, as hypothesized by our torch object analogy. In fact, the BFID*exp* model detected the bottle cap in ~60% of the images, compared to ~54% for the BFID model. Interestingly, both models sometimes detected the soother object which shares very similar features to a nursing bottle cap (~7% for BFID*exp* and ~11% of images for BFID). In many cases, the BFID model still detected the soother, while BFID*exp* model learned to detect the bottle cap instead. Some of these examples are illustrated in Fig. 5. This shows how data expansion can further teach our classifier to detect the correct object among two very similar ones. Spatially within the image, on average the BFID*exp* model detected an object at 111 pixels in distance to the nursing bottle while the BFID model detected the bottle at 126 pixels. This averaged 16-pixel difference may seem small but it was observed with

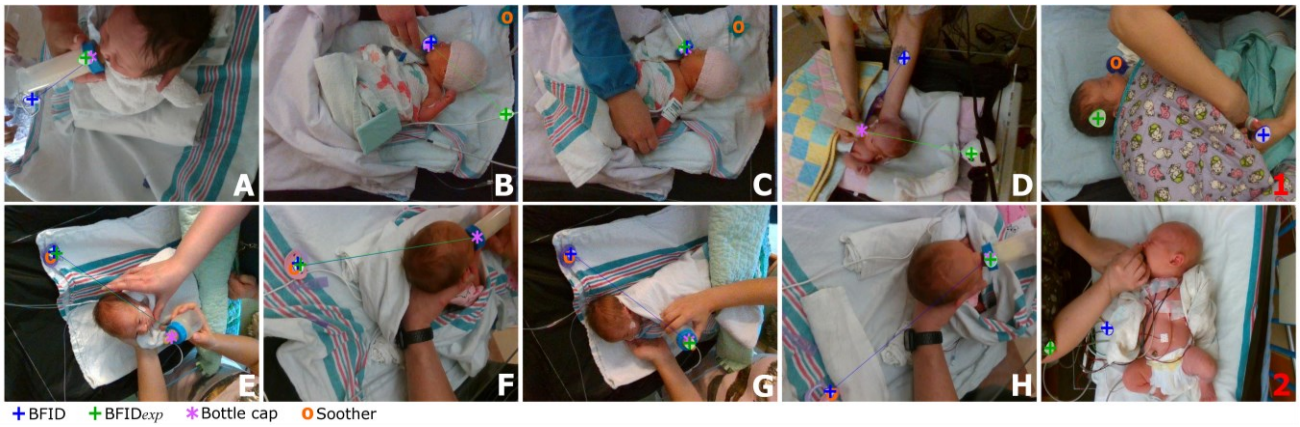


Fig.5 Domain Distance Mapping. Better performing model is A) BFID*exp*, B) BFID, C) both, and D) neither. The soother is detected in E-F) by both models, G-H) in BFID while BFID*exp* detects the bottle cap. 1-2) In absence of a bottle, both models detect other objects (ex: patient, nurse, cables).

a maximum of 254 pixels when the BFID model is closer, and 511 pixels for the BFID*exp* model (over twice as close). Our data expansion technique thus positively influences our model, given the closer distance to the bottle object. Other detections would include the patient's or nurse's arm or hand, the patient's head, toy, blankets, or cables, with comparable results from both models, and in absence of the bottle.

IV. Discussion

Overall, we obtained promising findings from our similar-feature data expansion method. When applying this technique, it is important to exclude the supplemented data in the evaluation since it could systematically learn to distinguish between one's own dataset and the outside sources. Future work will transition from classification to object detection, to further analyze the entire clinical scene (e.g., within bottle-feeding event, can we identify periods of active feeding vs. pauses). Our dataset solely included intervention images where the patient and nurse can be detected in the scene, and the occurrence of the nursing bottle would distinguish between the "bottle-feeding" and "no-feeding" class. Other combinations could be investigated for detailed analysis (e.g., patient present and bottle near the baby but absent nurse could suggest a paused feeding event). This might require more complex video analyses such as action recognition techniques. Other intervention events such as dressing, diaper change, or changing sensors will likewise require an evaluation of a sequence of images to infer context. Although, it may be difficult to find videos from outside sources to apply our data expansion technique, the approach may still be applicable since video analysis models often leverage a feature extraction step trained using individual images before concatenating frames to identify patterns in video sequences.

Our data expansion method can provide content but not always context. For example, an image of an adult holding a baby in one hand and a beer bottle in the other could satisfy our inclusion criteria (bottle & baby). Likewise, a photo of a child playing with empty beer bottles. However, these images have different meaning than a nursing baby. The original data collected dataset remains important to gain context for the classification task, while our similar-feature data expansion technique adds sufficient relevant content to address data scarcity and class imbalance. Not only is it time and cost effective compared to collecting new data, but it can substantially improve results for a difficult context-rich classification task. Ultimately, a future deployment of our model could improve patient care by assisting nurses in oral feeding assessments and documenting bottle-feeding events.

References

[1] M. Villarroel *et al.*, "Non-contact physiological monitoring of preterm infants in the Neonatal Intensive Care Unit," *npj Digit. Med.*, vol. 2, no. 1, pp. 1–18, Dec. 2019.

[2] S. L. Rossol *et al.*, "Non-Contact Video-Based Neonatal Respiratory Monitoring," *Children*, vol. 7, no. 10, p. 171, Oct. 2020.

[3] Y. S. Dosso, S. Aziz, S. Nizami, K. Greenwood, J. Harrold, and J. R. Green, "Video-Based Neonatal Motion Detection," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, vol. 2020-July, pp. 6135–6138.

[4] S. Orlandi *et al.*, "Detection of Atypical and Typical Infant Movements using Computer-based Video Analysis," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 3598–3601.

[5] M. S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, and Y. Sun, "Multi-channel neural network for assessing neonatal pain from videos," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2019, vol. 2019-Octob, pp. 1551–1556.

[6] Y. S. Dosso, S. Aziz, S. Nizami, K. Greenwood, J. Harrold, and J. R. . Green, "Neonatal Face Tracking for Non-Contact Continuous Patient Monitoring," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2020, pp. 1–6.

[7] R. Weber, A. Simon, F. Poree, and G. Carrault, "Deep transfer learning for video-based detection of newborn presence in incubator," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, vol. 2020-July, pp. 2147–2150.

[8] R. C. Cartwright-Vanzant, "Medical record documentation: Legal aspects in neonatal nursing," *Newborn Infant Nurs. Rev.*, vol. 10, no. 3, pp. 134–137, Sep. 2010.

[9] R. R. Hill, J. Park, and B. F. Pados, "Bottle-Feeding Challenges in Preterm-Born Infants in the First 7 Months of Life," *Glob. Pediatr. Heal.*, vol. 7, p. 2333794X2095268, Jan. 2020.

[10] B. F. Pados, J. Park, H. Estrem, and A. Awotwi, "Assessment tools for evaluation of oral feeding in infants younger than 6 months," *Adv. Neonatal Care*, vol. 16, no. 2, pp. 143–150, Apr. 2016.

[11] C. Lau, "Development of infant oral feeding skills: what do we know?," *Am. J. Clin. Nutr.*, vol. 103, no. 2, pp. 616S-621S, Feb. 2016.

[12] M. M. Palmer, K. Crawley, and I. A. Blanco, "Neonatal Oral-Motor Assessment scale: a reliability study.," *J. Perinatol.*, vol. 13, no. 1, pp. 28–35, Jan. 1993.

[13] S. M. Thoyre, C. S. Shaker, and K. F. Pridham, "The early feeding skills assessment for preterm infants.," *Neonatal network : NN*, vol. 24, no. 3. NIH Public Access, pp. 7–16, 2005.

[14] B. F. Pados, H. H. Estrem, S. M. Thoyre, J. Park, and C. McComish, "The Neonatal Eating Assessment Tool: Development and Content Validation," *Neonatal Netw.*, vol. 36, no. 6, pp. 359–367, Nov. 2017.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *The International Conference on Learning Representations*, 2015.

[16] L. Y. Pratt, J. Mostow, and C. A. Kamm, "Direct transfer of learned information among neural networks," in *AAAI'91 Proceedings of the ninth National conference on Artificial intelligence*, 1991, pp. 584–589.

[17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.

[19] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable Object Detection using Deep Neural Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2147–2154.

[20] L. Herranz, S. Jiang, and X. Li, "Scene recognition with CNNs: objects, scales and dataset bias," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[21] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis*, vol. 115, pp. 211–252, 2015.

[22] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *arXiv*, Dec. 2017.

[23] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv*, Nov. 2016.

[24] "Find your inspiration. | Flickr." .

[25] G. A. Miller, "WordNet: A Lexical Database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.

[26] "Home | Compute Canada." .

[27] X. Guo, W. Chen, and J. Yin, "A simple approach for unsupervised domain adaptation," in *Proceedings - International Conference on Pattern Recognition*, 2016, vol. 0, pp. 1566–1570.

[28] H. Wu, Y. Yan, M. K. Ng, and Q. Wu, "Domain-attention Conditional Wasserstein Distance for Multi-source Domain Adaptation," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 4, pp. 1–19, Jul. 2020.

[29] A. Bekele, J. Samuel, S. Nizami, A. Basharat, R. Giffen, and J. R. Green, "Ontology driven temporal event annotator mHealth application framework," in *28th Annual International Conference on Computer Science and Software Engineering*. IBM Corp., pp. 309–314, 2018.

[30] Intel, "Intel® RealSense™ Camera SR300 | Intel® Software." .