

Received June 9, 2021, accepted July 9, 2021, date of publication July 26, 2021, date of current version August 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099175

Cancer Registry Coding via Hybrid Neural Symbolic Systems in the Cross-Hospital Setting

HONG-JIE DAI^{1,2,3}, (Member, IEEE), YI-HSIN YANG², TI-HAO WANG⁴, YAN-JIE LIN⁵, PIN-JOU LU¹, CHUNG-YANG WU¹, YU-CHENG CHANG¹, YOU-QIAN LEE¹, YOU-CHEN ZHANG¹, YUAN-CHI HSU¹, HAN-HSIANG WU¹, CHENG-RONG KE¹, CHIH-JEN HUANG⁶, YU-TSANG WANG⁷, SHEAU-FANG YANG⁸, KUAN-CHUNG HSIAO², KO-JIUNN LIU⁹, LI-TZONG CHEN^{2,9,10}, I-SHOU CHANG^{5,11,12}, K. S. CLIFFORD CHAO¹³, AND TSANG-WU LIU^{11,12}

¹Intelligent System Laboratory, Department of Electrical Engineering, College of Electrical Engineering and Computer Science, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

²National Institute of Cancer Research, National Health Research Institutes, Tainan 70456, Taiwan

³School of Post-Baccalaureate Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

⁴Department of Radiation Oncology, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan

⁵Institute of Population Health Sciences, National Health Research Institutes, Miaoli 35053, Taiwan

⁶Cancer Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

⁷Division of Medical Statistics and Bioinformatics, Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

⁸Department of Pathology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

⁹Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung 80756, Taiwan

¹⁰Department of Oncology, National Cheng Kung University Hospital, National Cheng Kung University, Tainan 70403, Taiwan

¹¹National Institute of Cancer Research, National Health Research Institutes, Miaoli 35053, Taiwan

¹²National Institute of Cancer Research, National Health Research Institutes, Taipei 11503, Taiwan

¹³Cancer Center, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan

Corresponding authors: Hong-Jie Dai (hjdai@nkust.edu.tw), I-Shou Chang (ischang@nhri.edu.tw), K. S. Clifford Chao (d94032@mail.cmuh.org.tw), and Tsang-Wu Liu (walter@nhri.edu.tw)

This work was supported in part by the Ministry of Health and Welfare under Grant MOHW109-TDU-B-212-134026, and in part by the Ministry of Science and Technology of Taiwan under Grant MOST109-2221-E-992-074-MY3.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Kaohsiung Medical University Hospital under Application No. KMHIRB-EXEMPT (I)-20190216 and China Medical University Hospital under Application No. CMUH107-REC3-047, and performed in line with the Declaration of Helsinki.

ABSTRACT Cancer registries are critical databases for cancer research whose maintenance requires various types of domain knowledge with labor-intensive data curation. In order to facilitate the curation process with high quality in a timely manner, we developed a hybrid neural symbolic system for cancer registry coding. Unlike previous works which mainly worked on the dataset collected from one hospital or formulated the task as text classification problems, we collaborated with two medical centers in Taiwan to compile a cross-hospital corpus and applied neural networks to extract cancer registry variables described in unstructured pathology reports along with expert systems for generating registry coding. We conducted experiments to study the feasibility of the proposed hybrid for the task of cancer registry coding and compare its performance with state-of-the-art non-hybrid approaches. Furthermore, cross-hospital experiments were performed to study the advantages and limitations of transfer learning for processing reports from different sources. The experiment results demonstrated that the proposed hybrid neural symbolic system is a robust approach which works well across hospitals and outperformed classification-based baselines by F-scores of 0.13~0.27. Compared to the baseline models, the F-scores of the proposed approaches are apparently higher when fewer training instances were used. All methods benefited from the transferred parameters learned from the source dataset, but the results suggest that it is a better strategy to transfer the learned knowledge through the concept recognition task followed by the symbolic expert system to address the task of cancer registry coding.

INDEX TERMS Electronic medical records, medical expert systems, medical information systems, natural language processing.

The associate editor coordinating the review of this manuscript and approving it for publication was K. Kotecha¹.

I. INTRODUCTION

Cancer is a main cause of mortality worldwide and has been the leading cause of death over several decades in Taiwan.

The Taiwan Cancer Registry (TCR), a population-based cancer registry, was established by the Taiwan Society of Cancer Registry in 1979 and was supported by the Ministry of Health and Welfare (MOHW) over 40 years to provide a comprehensive measurement of cancer incidence, morbidity, survival, and mortality for persons with cancer in Taiwan. Hospitals with greater than 50-bed capacity providing outpatient and hospitalized cancer care are recruited to participate in reporting all newly diagnosed malignant neoplasms to the registry. Full-time staffs with expertise of various fields such as pathology, clinical oncology, radiation oncology, and cancer registrar are involved in the process of manually drafting and examining medical and pathology records to determine the eligibility of the patient for cancer registry.

Delay time reporting refers to the time that elapsed before reporting a diagnosed cancer case to the cancer registry. In both Taiwan and United States, the delay time between the time of cancer diagnosis and the time of publicly reporting the cancer incidence data is about 2 years [1]. A delay in the reporting of cancer cases results in an underestimation of the cancer rate in a population. Unfortunately, the process of reporting cancer cases requires manual review of large volumes of reports such as pathology reports and radiology reports, which is pronouncedly labor-intensive and time-consuming.

One solution to this problem currently being explored is the application of artificial intelligence (AI) techniques to automatically read and extract information from cancer pathology reports. However, as cancer pathology reports are often represented in unstructured format, it can be difficult for machines to explicitly process them due to the diverse writing styles among different hospitals. Therefore, an efficient and precise approach is critical in reducing the overall effort to accelerate the reporting of cancer cases, which also allows trained personnel to focus on cancer related analysis and researches.

To improve the collaboration and resource sharing of hospitals and cancer research centers in Taiwan, a national project named the Integrated Cancer Research Multicenter Collaboration Platform (ICRMCP) was established under the cancer center support grant program funded by the MOHW. One major goal of ICRMCP is to apply natural language processing (NLP) techniques to address the issue of delayed reporting and further provide a cancer research information sharing service to promote the sharing of cancer research patient information and computing resources from participating hospitals.

In this study, we introduce our approach to automatically extract eight primary cancer registry codes from unstructured pathology reports. Unlike previous similar works [2]–[5] which mainly formulated the task as text classification problems and applied a variety of machine learning methods to develop black box nature models for cancer registry variables, we presented a hybrid strategy which combines a deep sequential labeling neural network with an expert system to extract high quality cancer registry coding informa-

tion and also provide explainability. This method mimics the duty of a cancer registrar, which includes analyzing unstructured reports to identify cancer registry items and translating these items into standardized codes. Furthermore, we employed transfer learning and conducted experiments to examine the performance of the proposed approaches on the cross-hospital pathology materials to gain insights on the effectiveness of transfer learning and the robustness of the developed knowledge rules.

II. METHODS

A. TASK DEFINITION AND DATASETS

In the first year of the ICRMCP project, we primarily focused on the analysis of the pathology reports for colorectal cancer, which is the third leading cause of cancer-specific deaths in Taiwan. We collaborated with two medical centers, namely China Medical University Hospital (CMUH) and Kaohsiung Medical University Chung-Ho Memorial Hospital (KMUH), for the collection of patients' cancer registry records linked with the corresponding colorectal pathology reports for generating these records. Non-tumor reports and reports without corresponding cancer registration records were excluded. Fig. 1 shows a snippet of a collected pathology report. Although the official language in Taiwan is Mandarin Chinese, all collected pathology reports were written in English.

Unlike pioneering projects whose goal is to develop general purpose NLP systems for extracting cancer-related phenotypes [6], [7] or populating knowledge representation models [8], we focused on the construction of a NLP tool to facilitate cancer registration. In this pilot study, we examined the TCR coding manual¹ and the American Joint Committee on Cancer (AJCC) staging manual 8th version [9] to select eight coding items including pathological TNM classifications (TNM), the number of examined nodes (EN) and positive nodes (PN), tumor size (TS), histology types (H), and grades (G). The coding results associated with a registry record curated by the registrar were collected in our corpus. Note that a registry record could refer to more than one pathology report.

Corresponding to the eight coding items, we further defined nine cancer registry concepts which should be extracted for making the judgement of coding in our annotation guideline created by consulting a committee composed of hospital investigators and cancer registrars. The concepts include descriptions related to the aforementioned eight items and the stage classification (SC). The entire annotation process for the nine concepts was described in our previous work [10]. Note that we asked the annotators to annotate all of mentions of the nine concepts among all pathology reports associated with a cancer registry instead of only annotating the mentions which were used as the reference for coding.

In a nutshell, our corpus contains narrative pathology reports written in English with two types of annotations.

¹The manual is available at http://tcr.cph.ntu.edu.tw/uploadimages/Long-form%20Manual_Official%20version_20190304_W.pdf.

One is for the recognition of cancer registry concepts which were annotated by NLP annotators, and the other is for cancer registry coding curated by registrars. Based on the compiled corpus, two subtasks were defined:

1. The registry concept recognition subtask: The goal of this subtask is to recognize registry concepts described in a report. The text spans surrounded by brackets shown in Fig. 1 indicate the annotations for cancer registry concepts. The subscript preceding the right bracket indicates the concept type. For example, “3.7” and “6.5” are annotated with TS indicating the size of tumor. A total of seven concepts were annotated in Fig. 1.

... Microscopically, it shows a [moderately-differentiated_G] [adenocarcinoma_H] of rectum (tumor size about [3.7_{TS}] cm in length and [6.5_{TS}] cm in width) arising in the background of [tubulovillous adenoma_H] with severe dysplasia, invading deeply into perirectal soft tissue. The outermost surface of perirectal soft tissue is focally very close to tumor (less than 0.5 mm in distance). ... All of the [21_{EN}] dissected regional lymph nodes are negative of malignancy. The immunohistochemistry study of EGFR stain for primary tumor shows mild to moderate basolateral membrane staining in 30% of tumor cell. The [two_{EN}] lymph nodes in the specimen labeled as "LN" are negative of malignancy. ...

FIGURE 1. A snippet of a collected pathology report.

2. The cancer registry coding subtask: The aim of this subtask is to generate the coding items for the eight concept types including T, N, M, EN, PN, TS, H, and G. Take the snippet shown in Fig. 1 as an example. The codes curated for the snippet are EN: 23, G: 2, H: 82633, and TS: 065. Note that the code assigned for TS is 065 and not 037 because the coding manual requires only the maximal diameter to be coded. The code for EN is the sum of all examined nodes (21 + 2). It is important to note that this simply an example, and in real-world scenarios registrars should consider all available reports of a patient to curate the final codes.

B. HYBRID NEURAL SYMBOLIC SYSTEM FOR CANCER REGISTRY CODING

We followed the loosely coupled architecture [11] to develop a hybrid neural symbolic system with its architecture illustrated in Fig. 2. The system contains three layers in which a deep neural network is housed at the second layer for extracting cancer registry-related concepts (the first subtask) from preprocessed pathology reports from the first layer. At the last layer the symbolic expert system generates the final coding results (the second subtask). The entire coding process is elaborated as follows.

1) SYSTEM WORKFLOW

Before processing a new list of pathology reports of a patient, all updated facts in the database of our expert system are reset to the initialization state. Each pathology report is preprocessed by the first layer in which our clinical toolkit [12] was employed to segment sentences, generate the corresponding

tokens and recognize section headings, such as “Microscopic Examination”. The preprocessed sentences are then analyzed by our neural network to extract the nine registry concepts. The concepts along with the preprocessed information is established as new “Facts” in the “Fact Database” as shown in the layer 3 of Fig. 2. A fact here represents the observed coding clues from the report being processed. We will describe it in detail in the following subsection.

Starting with the known facts in the fact database, the “Inference Engine” of our expert system applies the forward-chaining algorithm to trigger all rules in the “Knowledge Base” whose antecedents are satisfied with the current facts and updates their conclusions to the fact database to form new known facts. This process is repeated until no additional rules can be fired, and the final coding results can be obtained from the system. Two expert systems were developed to accomplish the coding task. One is in charge of processing each individual pathology report in order, and the other considers the results from the first one to generate the final codes.

2) REGISTRY CONCEPT RECOGNITION LAYER

Several approaches have been proposed to extract cancer registry-related concepts based on the named entity recognition techniques [10], [13]–[16]. In the layer 2 we used our neural network [10] to recognize registry concepts mentioned in reports by applying the IOB (Inside-Outside-Beginning)-2 tag scheme to formulate the problem as a sequential labeling task. The input of the network is the pre-processed sequence of tokens in a pathology report with the output being the sequence of labels for each token. The input tokens were represented as a vector by concatenating the pre-trained word representations obtained by using GloVe [17] and RoBERTa [18]. The values of the concatenated vectors were fixed during the training process.

The concatenated representation was then fed to a fully connected layer (denoted as FC1 in Fig. 2) along with a variational dropout before passing the embeddings into the bidirectional long-short term memory (BiLSTM in Fig. 2) network with one layer consisting of 256 hidden nodes. The output of the BiLSTM layer goes through another fully connected layer (denoted as FC2 in Fig. 2) to generate an output of a size equal to the number of the labels. The output of FC2 is then inputted to the inference layer based on the conditional random field (CRF in Fig. 2) to model the dependencies between labels in neighborhoods with the Viterbi loss to jointly decode the best chain of labels for the given sequence.

3) CANCER REGISTRY CODING LAYER

The information extracted by the layers 1 and 2 is transformed to facts and stored in the fact database for the expert systems in the third layer to generate the corresponding codes.

Instead of representing facts as flat assertions about observed cancer registry-related information, they are represented by objects whose asserted values can be accessed by

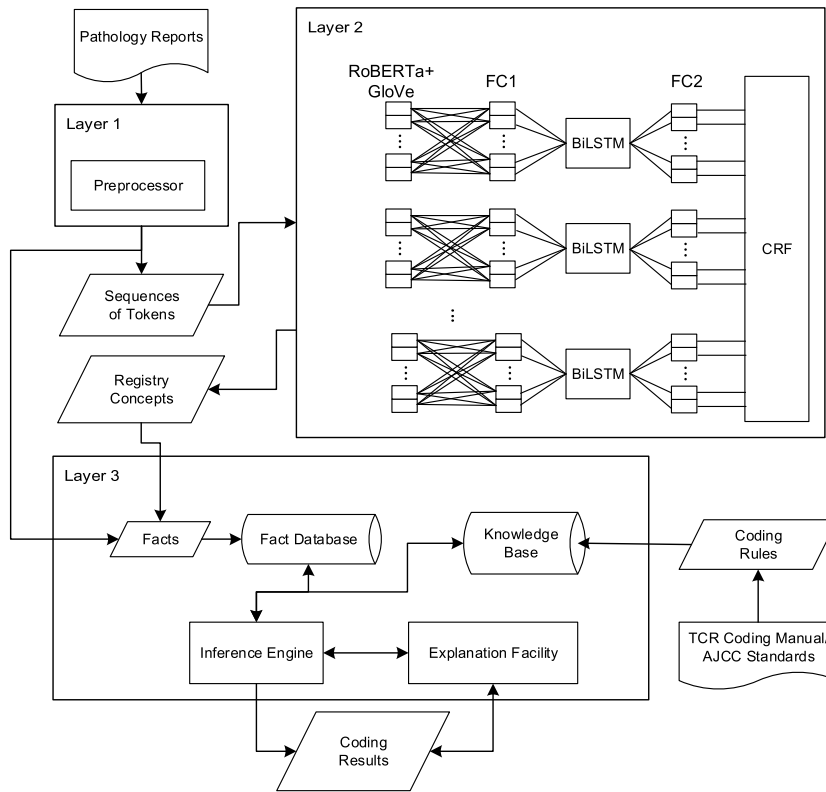


FIGURE 2. Global architecture and workflow of the developed hybrid neural symbolic system for cancer registry coding.

the dot (·) operator. Before storing the facts in the database, those with numeric values in word forms are normalized to integers. Take the two EN concepts shown in Fig. 1 as an example. Two fact objects, C_1 and C_2 , are stored in the database whose asserted values include:

1. *Type*: “EN” for both concepts.
2. *Value*: Integer values “21” and “2” for the first and the second EN concept, respectively.

The asserted type and value attributes of the C_1 object can be accessed by the syntaxes “ $C_1.type$ ” and “ $C_1.value$.”

To construct our expert systems, the knowledge engineers P.J.L. and H.J.D. interviewed two cancer registrars C.J.H. and S.F.Y. and followed the TCR coding manual (Long Form) and the 8th edition of the AJCC standards to encode the knowledge for cancer registry coding in our expert system. The acquired knowledge was represented by rules that were stored in the knowledge base (KB) shown in Fig. 2. Each rule consists of the antecedent and the consequent parts. The antecedent of a rule represents the desired condition to be satisfied for firing the rule. A rule can have multiple antecedents joined by the keyword AND. An antecedent is represented by a fact object with the desired value linked by an operator. For example, the antecedent below represents a condition in which the number of observed examined regional lymph nodes is less than 90.

$$C.type = EN \text{ AND } C.value < 90$$

The antecedent of a rule is matched against the observed facts stored in the fact database. On the other hand, the consequent of a rule indicates the action to be performed if the antecedent is satisfied. For instance, the antecedent in the following rule is satisfied if all available EN facts located in the “microscopic” section have the same “value” attribute. The subsequent action is then to update the database with a new fact D whose “type” and “value” attributes are set to “EN_CHECKED” and that of the first element in the list containing all available EN facts, respectively.

$$\begin{aligned} \text{AreSame}(\{C|C.type = EN \text{ AND } C.section \\ &= \text{'MICROSCOPIC'}\}, \text{'value'}) \\ &= \text{TRUE} \rightarrow D.type \\ &= EN_CHECKED \text{ AND } D.value \\ &= C_1.value \text{ AND } D.section \\ &= \text{'MICROSCOPIC'} \end{aligned}$$

In our implementation of the expert system developed for processing a single pathology report, the observed facts incorporate the initial facts added by the preceding layers and other facts dynamically added to the database during the forward reasoning process of our inference engine. The following rules are defined in the expert system to generate

code for the EN concepts shown in Fig. 1.

$$\begin{aligned} \text{AreSame}(\{C|C.type = \text{EN AND } C.section \\ &= \text{'MICROSCOPIC'}\}, \text{'value'}) \\ &= \text{FALSE AND IsSum}(\{C|C.type \\ &= \text{EN AND } C.section \\ &= \text{'MICROSCOPIC'}\}, \text{'value'}) \\ &= \emptyset \rightarrow D.type \\ &= \text{EN_CHECKED AND } D.value \\ &= \text{Sum}(\{C|C.type = \text{EN AND} \\ &C.section = \text{'MICROSCOPIC'}\}, \\ &\text{'value'}) \\ \text{Size}(\{C|C.type = \text{EN_CHECKED}\} = 1 \rightarrow D.type \\ &= \text{EN_SINGLE_CODING AND} \\ &D.value = C.value \end{aligned}$$

The first rule states that if the values of all recognized EN concepts located in the “microscopic” section are not the same and there is no EN concept whose value is the sum of the value of all remaining concepts, then the database is updated with a new fact D whose type and value attributes are “EN_CHECKED” and the sum of the values of all recognized EN concepts, respectively.

If the inference engine can derive only one “EN_CHECKED” fact in the fact database, the second rule is triggered and sets the EN code for the report to be the value of the “EN_CHECKED” fact. If there are more than one “EN_CHECKED” facts, we follow the guidance of the TCR coding manual to select the value of the derived “EN_CHECKED” fact in the following order prioritized by the section heading: “Final diagnosis” > “microscopic” > “gross”.

After individually processing all pathology reports for one patient, we have a list of codes assigned for each report ordered by date. The list is then stored in the fact database for the second expert system that was developed to consider all coding results among the available reports. In our current implementation, the rules developed for the second system is straightforward by considering the guidance of the TCR coding manual and the temporal information. Take the TS concept shown in Fig. 1 as an example. The first expert system outputs “065” as the code for the snippet. The second expert system uses the following rule to output “065” as the final code if the value is the maximal diameter among all reports.

$$\begin{aligned} \text{Size}(\text{Remove}(\{C|C.type = \text{TS_SINGLE_CODING}\}, \\ \text{'value'} = \{990, 991, 992, 993, 994, 995, 998, 999\}) \text{AS } D) \\ > 0 \rightarrow E.type = \text{TS_CODING AND } E.value \\ &= \text{Max}(D, \text{'value'}) \end{aligned}$$

The rule above removes TN_SINGLE_CODING facts with values in the range of 990 to 999 prior to identifying the maximum diameter value since these values have unique

definitions. For instance, “990” indicates “no size of focus is given”, and “991” indicates the size is less than 1 cm.

C. METHODS DEVELOPED FOR PERFORMANCE COMPARISON

1) BASELINE MODELS

Gao *et al.* [2] proposed to use the hierarchical attention network (HAN) [19] to extract primary sites and G from pathology reports and demonstrated that HAN significantly outperformed the conventional machine learning and deep learning techniques. Therefore, we accommodated their work as a baseline system for performance comparison of the developed hybrid neural symbolic system. We followed the binary relevance transformation method [20], [21] to transform the original problem into eight binary classification tasks and learned eight corresponding HAN classifiers, with one for each registry variable. The same concatenated vectors consisting of GloVe and RoBERTa were used to represent words.

Previous works [10], [22] have demonstrated that CRF and neural network-based approaches outperformed traditional methods based on dictionaries and support vector machines for the first subtask. In this study, we implemented the CRF model with features including the cluster-based word representation feature where the number of clusters was set to 1,000 [23] based on GloVe, and the part-of-speech and normalized unigram and bigram features [24] within a context window of five along with transition features for the first subtask. We then replaced the BiLSTM-CRF model in the second layer with the developed CRF model to develop another baseline for comparison. The expert systems along with the same rule sets were used in the third layer.

Furthermore, considering transformer-based language models have taken over several NLP tasks [25] and have been applied by top-performing teams in shared tasks for clinical NLP [26], we followed the end-to-end learning approach suggested by Devlin *et al.* [25] to directly address the cancer registry coding task. We fine-tuned RoBERTa on our corpus by adding an additional multilayer perceptron (MLP) whose input is the final hidden vector generated by RoBERTa corresponding to the first input token. All pathology reports were truncated to 512 tokens to meet the requirement of RoBERTa.

D. TRANSFER LEARNING FROM A SOURCE HOSPITAL TO A TARGET HOSPITAL

Transfer learning [27] aims to learn a better model on a target domain by leveraging the knowledge previously learned from a source domain. In this study, we treated KMHU as the source hospital and CMUH as the target hospital because the size of the KMHU dataset is larger than that of the CMUH dataset. The inductive transfer learning technology was applied for transferring the learned parameters from the model of KMHU.

For the developed BiLSTM-CRF neural network, we conducted experiments to study the effectiveness of the transfer learning for the recognition task by transferring different

numbers of the learned parameters of the developed models to the target hospital by retraining the model with transferred parameters on the target hospital’s dataset via fine-tuning. We then hybridized the transferred models with the developed symbolic expert system to study the effect of the transferred model for the entire hybrid neural symbolic system.

As for the HAN models, we replaced the final MLP layers of the KMHU-HAN model with a new MLP whose dimension equals to the number of possible unique labels for a registry item observed in the CMUH dataset. Note that as shown in Fig. 3, the number of possible values for a registry item can be different in the two hospitals. Whereas for the pre-trained RoBERTa model, we applied twice fine-tuning strategy [28] to fine-tune KMHU-RoBERTa on the CMUH dataset; the output dimension of the added MLP was extended to meet the required numbers of unique labels for the CMUH dataset. The weights for the new dimensions were randomly initialized and the weights for the labels appeared in both hospitals were transferred.

E. EVALUATION METRICS AND EXPERIMENT CONFIGURATIONS

We conducted experiments to compare the performance of the developed hybrid symbolic neural system with baselines to investigate the effectiveness of the proposed systems and the effect of the transfer learning for the developed systems and baseline models. The standard macro-precision (P, *a.k.a.* positive predictive value), recall (R, *a.k.a.* sensitivity) and F-measure (F) defined as follows were used to evaluate the performance of the methods developed for both the recognition and coding tasks.

$$Precision_i / \text{Positive Predictive Value}_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$\text{Recall}_i / \text{Sensitivity}_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$\text{Macro - P} = \text{average(Per - class P)} = \frac{\sum_i P_i}{\text{Number of Classes}} \quad (3)$$

$$\text{Macro - R} = \text{average(Per - class R)} = \frac{\sum_i R_i}{\text{Number of Classes}} \quad (4)$$

$$\text{Macro - F - measure} = 2 \frac{\text{Macro-P} \times \text{Macro - R}}{\text{Macro-P} + \text{Macro - R}} \quad (5)$$

In the formulae above, TP_i , FP_i , and FN_i represent the number of true positives (TPs), false positives (FPs), and false negatives (FNs) for the registry item type i , respectively. The TPs, FP and FN are defined on the token-level for subtask 1, whereas they are defined on the document-level for subtask 2.

To train the neural networks in our experiments as well as to examine the knowledge represented in the form of production rules of the expert system, we randomly kept 50 reports in the training sets as the validation sets which were excluded from training to determine the best-performing models and

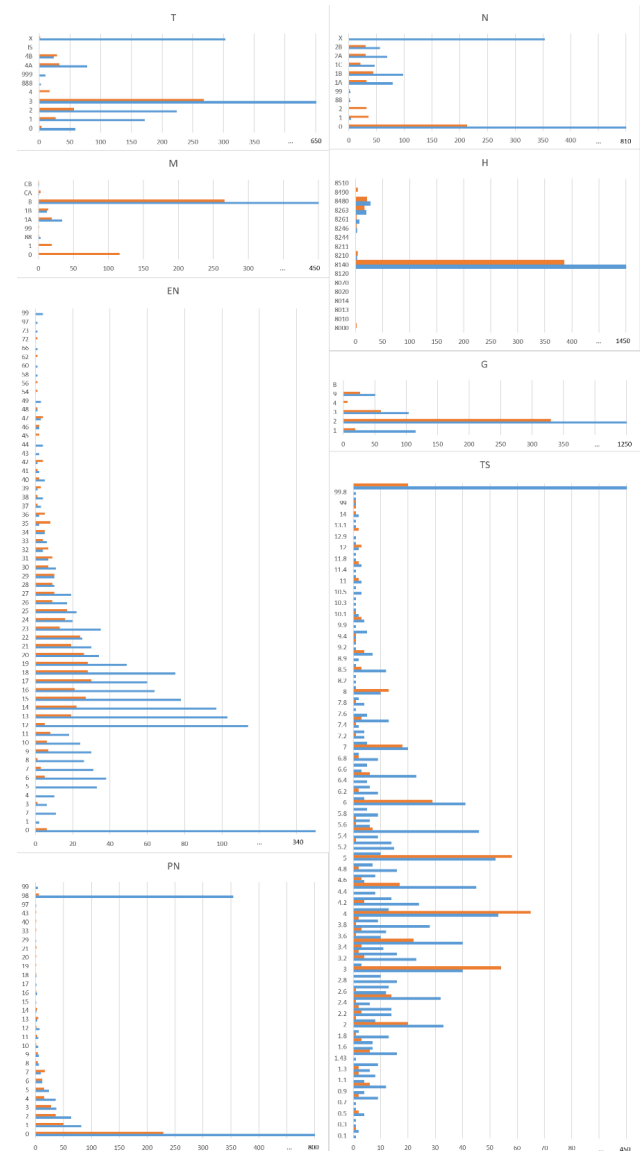


FIGURE 3. Statistics of the datasets collected. The blue bars and the orange bars represent the categories of the KMHU and CMUH datasets, respectively.

improve the comprehensiveness of the developed rules in the development phase.

III. RESULTS AND DISCUSSION

A. CORPUS STATISTICS

We collaborated with CMUH and KMHU who provided 393 and 2,347 reports, respectively, to assemble our dataset along with the annotations for registry concepts and coding results curated by registrars based on the reports that correspond to the treatment trajectory of the patient. Table 1 displays a statistical summary of the compiled corpora for the subtask 1. Note that in the KMHU corpus, a cancer registry record of a patient may associate with more than one pathology reports. A record has at least one and at most four records in the KMHU dataset.

For the subtask 2, eight registry items (T, N, M, EN, PN, TS, H, and G) were extracted from the curated coding results

TABLE 1. Corpus statistics for the compiled corpora for the registry concept recognition subtask.

Item ^a	CMUH			KMUH		
	Training	Test	Total	Training	Test	Total
T	400	122	522	1,440	99	1,539
N	380	122	502	898	61	959
M	373	124	497	6	0	6
TS	1,112	383	1,495	1,606	115	1,721
EN	623	189	812	2,021	153	2,174
PN	554	177	731	2,021	153	2,175
SC	670	161	831	1,441	99	1,540
H	558	136	694	4,517	273	4,790
G	519	140	659	4,410	265	4,675
NR	293	100	393	2,207	140	2,347
NP	293	100	393	1,515	100	1,615

^aT = pathological T, N = pathological N, M = pathological M; TS = tumor size, EN = numbers of examined nodes, PN = numbers of positive nodes, SC = staging classification, H = Histology, G = grade, NR = total number of reports, NP = total numbers of patients.

to create the gold annotations. These items were part of the long form data submitted to TCR [29]. Fig. 3 shows a statistical summary of the data collected for each registry item. In general, we observed that the coverage of the KMUH dataset is superior to that of the CMUH dataset, with the former particularly covering a much wider range of values of EN and PN. The values of EN in the KMUH and the CMUH datasets range from 0 to 99 and 0 to 72, respectively. The KMUH dataset contains at least one training instance for the EN values between 0 to 44, but there are six values that do not have any training instances in the CMUH dataset. Four samples in the KMUH dataset had an EN code of 99, indicating that the numbers of examined nodes are unknown for the patients.

On the other hand, excluding the coding values larger than 99, the distribution of the tumor size is a left-skewed normal distribution. The tumor sizes range from 0.1 cm to 15 cm with a mean of 4.44 cm. The smallest and largest tumor sizes in the greatest dimension observed in KMUH and CMUH are 0.1 cm and 14 cm and 0.2 cm and 15 cm, respectively. The CMUH dataset only contains a wider range of values in the category of the pathological M classification. For other categories such as G and PN, both datasets had a similar range of values, but the KMUH dataset carries more training instances for each coding value owing to its larger size.

B. PERFORMANCE COMPARISON OF CANCER REGISTRY RECOGNITION AND CODING

In the experiment, we first trained the developed CRF and BiLSTM-CRF models on the two training sets separately and evaluated their recognition performance on the test sets of the two hospitals. We then examined the effectiveness of the developed expert system for cancer registry coding by hybridized it with the two models and compare their performance with the baselines including HAN and RoBERTa. The results are depicted in Table 2.

In general, for both the recognition and coding tasks, all the developed methods including the proposed hybrids,

baselines and transformer-based models performed better on the KMUH test set which may be due to the larger number of training samples. The hybrid neural symbolic systems achieved the highest overall F-scores on the dataset of both hospitals.

1) RESULTS OF REGISTRY CONCEPT RECOGNITION

Because both HAN and RoBERTa were implemented by formulating the task as text classification problems, we cannot report their recognition performance in Table 2. For the recognition subtask, CRF-S achieved a comparable overall F-score on the KMUH test set, but its F-score is lower than that of BiLSTM-CRF-S on the CMUH test set owing to its poor recall of 0.423 on the recognition of TS mentions. In accordance with the observation of Wallace *et al.* [30], we noticed that the inclusion of the pre-trained word representation features enables the CRF model to learn the numeracy for integers and their corresponding word forms (*e.g.* 42 and forty-two). After clustering, we found that the word forms of the integer values less than ten belong to the same cluster of the integers whose values are less than 30. The number magnitude is also captured by the different cluster numbers assigned to the different continuous ranges of integers. However, the float values and their corresponding word forms (*e.g.* 1.6 and one point six) cannot be accurately identified by the developed clusters leading to the significantly lower recall of the model. On the contrary, the neural network-based models can directly exploit the pre-trained word representations so that outperformed the CRF model. We didn't find the cases of word forms in the KMUH dataset, so that both methods achieved satisfied F-scores.

2) RESULTS OF CANCER REGISTRY CODING

For the coding task, we can see that the coding performance of the hybrid systems is lower than their recognition performance but both CRF-S and BiLSTM-CRF-S achieved higher overall F-scores than that of the other two classification-based models (HAN and RoBERTa). Unlike the baseline models learned in the end-to-end manner, the recognition performance of the developed hybrid system indeed affects the coding performance in case that the performance of the recognition is apparently lower, such as TS or H in CMUH. In general, the impact of different factors on the coding performance is subtle but somewhat depends on the characteristics of the corresponding registry item. For instance, in Table 1 we can see that TS items are frequently mentioned in the pathology reports provided by CMUH, and the guideline requires curators to examine all mentioned TS items to make their decisions. The R rate of the TS recognition is therefore more important than P.

Although the KMUH corpus is much larger than the CMUH corpus, the latter had more annotations for the variable pathological M. This is because the judgement of the pathological M stage requires additional reports such as image reports from other examination divisions, and consequently the conclusion of pathological M stage often remains

TABLE 2. The comparison of the macro-F-scores of different methods on the test sets for the task of concept recognition and coding.

Registry Concept/Item	CMUH						KMUH					
	CRF-S ^a		BiLSTM-CRF-S ^b		HAN ^c	RoBERTa ^d	CRF-S ^a		BiLSTM-CRF-S ^b		HAN ^c	RoBERTa ^d
	R ^e	C ^f	R ^e	C ^f	C ^f	C ^f	R ^e	C ^f	R ^e	C ^f	C ^f	C ^f
T	0.962	0.929	0.996	0.939	0.930	0.770	0.995	0.881	0.980	0.883	0.877	0.890
N	0.929	0.940	0.918	0.960	0.770	0.820	0.992	0.895	0.976	0.895	0.930	0.890
M	0.949	0.842	0.975	0.853	0.780	0.800	n/a	1.000	n/a	1.000	1.000	1.000
TS	0.564	0.700	0.931	0.850	0.150	0.140	0.925	0.719	0.974	0.860	0.263	0.180
EN	0.951	0.900	0.946	0.910	0.070	0.610	0.993	0.860	0.980	0.860	0.368	0.160
PN	0.963	0.945	0.960	0.960	0.600	0.670	0.980	0.893	0.987	0.893	0.930	0.910
SC	0.968	n/a	0.964	n/a	n/a	n/a	0.995	n/a	0.990	n/a	n/a	n/a
H	0.903	0.902	0.778	0.856	0.780	0.850	0.991	0.955	0.984	0.955	0.930	0.947
G	0.924	0.870	0.922	0.885	0.780	0.920	0.998	0.895	1.000	0.930	0.930	0.860
Overall	0.908	0.879	0.934	0.903	0.608	0.698	0.984	0.888	0.984	0.910	0.779	0.730

^aCRF-S = hybrid of conditional random fields-symbolic expert system.

^bBiLSTM-CRF-S = hybrid of bidirectional long-short term memory-conditional random fields-symbolic expert system.

^cHAN = hierarchical attention network.

^dRoBERTa = robustly optimized bidirectional encoder representations transformers approach.

^eR = recognition task (subtask 1).

^fC = coding task (subtask 2).

undecided based on the current pathological data in the KMUH corpus. Based on the TCR coding manual, most registry items considered in this study have default coding in cases where the reports did not contain sufficient information to determine the code of a target registry item. Therefore, even though the M concept is absent in the KMUH test set as shown in Table 1, our expert system follows the manual to assign the code “999” to the reports. Both classification-based methods also learned to assign the correct coding for these reports.

Classification-based methods can achieve comparable F-scores for some variables on the KMUH dataset but had an apparently worse performance on CMUH, which may be owing to the small size of the training set. The superior performance of RoBERTa to HAN on the CMUH test set demonstrates the effectiveness of the pre-trained language model when less labeled data was provided. On the other hand, the HAN model outperformed RoBERTa on the KMUH corpus possibly owing to the length of the KMUH reports, which is much longer than that of the CMUH reports. A KMUH patient has an average of 1.46 reports (refer to Table 1), whose content may exceed the limitation of 512 tokens of RoBERTa. In this case, the hierarchical-oriented architecture like HAN with pre-trained RoBERTa as feature extractor may be more competent. Unfortunately, both models tend to obtain worse performance in items with numeric values as their codes.

C. EFFECTIVENESS OF TRANSFER LEARNING FOR THE RECOGNITION TASK AND THEIR EFFECT OF THE CODING TASK

In the following experiments, we would like to gain insight into the extent of improvement transfer learning brings on the performance of cancer registry recognition and their effect in the coding task on the cross-hospital setting. KMUH was selected as the source dataset since its larger size. The results of for the recognition and coding are respectively illustrated in the sub-graphs (a) and (b) of Fig. 4.

For all the configurations prefixed with “Non-transferred” shown in Fig. 4, we only used the reduced sizes of

the CMUH training set to train the neural network models without relying on any pre-trained parameters on the KMUH dataset. The following four configurations were considered.

1. Non-transferred BiLSTM-CRF (for recognition)
2. Non-transferred BiLSTM-CRF Hybrid (for coding)
3. Non-transferred HAN (for coding)
4. Non-transferred RoBERTa (for coding)

Furthermore, considering the comparable results achieved by the CRF models, we also included the configurations “Non-transferable-CRF” and “Non-transferable-CRF Hybrid” in which we trained six CRF models corresponding to the reduced CMUH datasets and hybrid them with the expert system with the same rule sets.

1) EFFECTIVENESS OF TRANSFER LEARNING FOR THE RECOGNITION TASK

For the recognition task, we conducted experiments to examine the effect of transfer knowledge by analyzing the importance of each layer of the developed neural networks and quantifying the performance gain by varying the sizes (5%, 20%~100%) of the CMUH training set when we fine-tuned the model pre-trained on KMUH. The “FC1” configuration in the sub-figure (a) initialized the learned parameters of the FC1 layer of the BiLSTM-CRF model by adopting the pre-trained parameters on KMUH, while “BiLSTM” further included the learned parameters of the BiLSTM layer and so on.

From the results illustrated in the sub-figure (a) of Fig. 4, we can observe that even with only 20% of the CMUH training set, models learned with the transferred parameters achieved promising F-scores of at least 0.902, which outperformed the “Non-transfer BiLSTM-CRF” and the “Non-transfer-CRF” models trained on the same number of training samples by 0.241 and 0.116, respectively. Performance can be further improved for both transferred and non-transferred models with increased numbers of training samples, while the latter exhibited a more apparent performance enhancement. But the transferred models trained on the 20% training

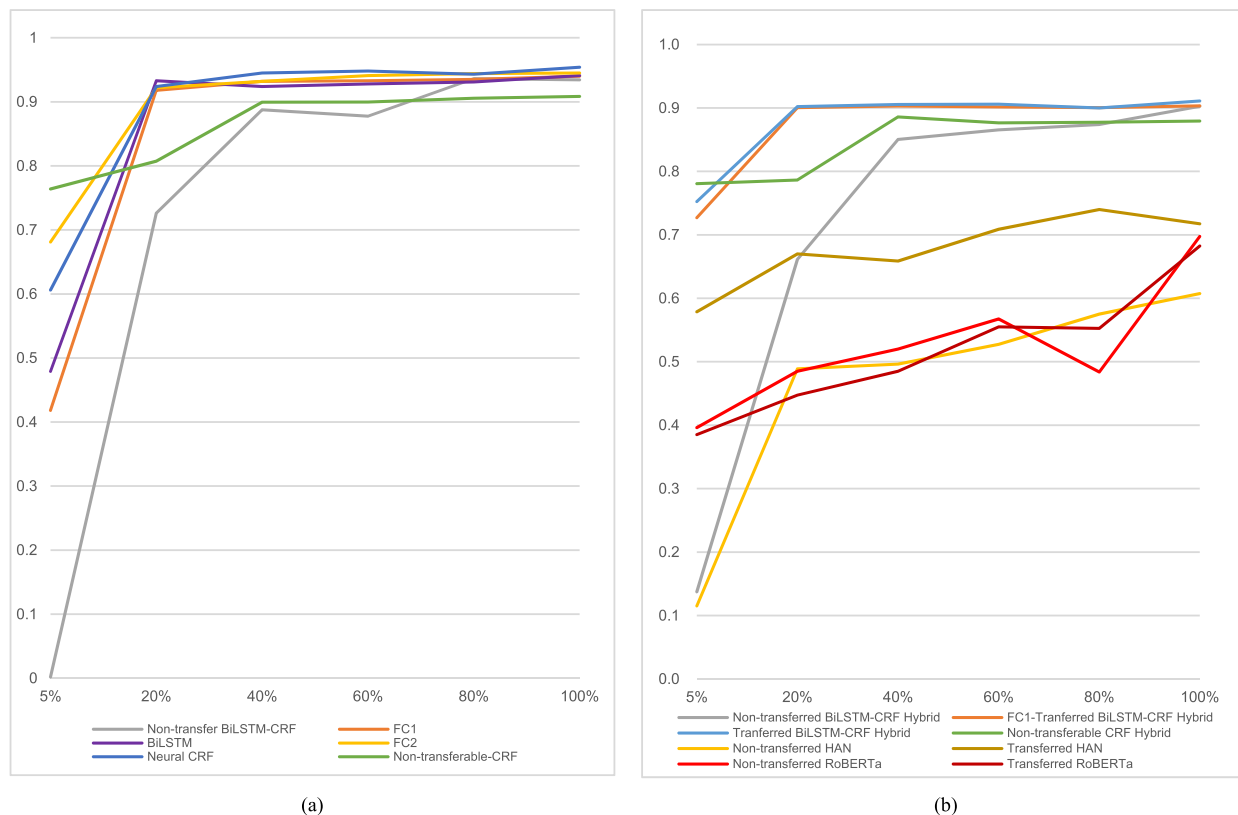


FIGURE 4. (a) Impact of the overall macro avg. F-scores by fine-tuning the models with the parameters up to each layer pre-trained on KMUH on the varied sizes of the CMUH training set. (b) Impact of applying transfer learning in the recognition task for cancer registry coding by using the hybrid of the pre-trained KMUH models fine-tuned on various sizes of the CMUH training set and the symbolic expert system.

set still obtained competitive F-scores that outperformed all non-transferable CRF models. The improvement of the CRF models is relatively flat compared to that of the neural networks, but it outperformed the others when the 5% training set was used, highlighting the importance of handcraft features when only very few training instances were available.

The results also reveal the effectiveness of the learned parameters of each layer of the developed model in the manner of transfer learning. It is worth noting that when very few training instances were provided, transferring parameters of all layers to the target model seems to be a good strategy, which in general led to slightly better F-scores. But we also observed that transferring the parameters of the lower layers is almost as efficient as transferring all parameters in cases that enough training samples were provided considering their high performance when we trained them on more than 20% target training set. This phenomenon is consistent with the observations of other previous works [31], [32] and the hypothesis that the lower layers of a neural network learn generic features and the higher layers learn task-specific (in this work hospital-specific) features.

2) THE EFFECT OF TRANSFER LEARNING FOR THE CODING TASK

In this experiment, we investigated the effect of applying transfer learning to the task of cancer registry coding. Since we implemented both the proposed hybrid neural

symbolic approach and the classification-based approach to address the coding task, we examined the effect of transfer learning from two perspectives including the benefit of transfer learning for classification-based approaches, and the influence of transferred neural network on the developed symbolic expert system. Similar to the experiment conducted in the previous subsection, we considered the following configurations for comparison with the “non-transferred/transferable” configurations by training or fine-tuning them on various CMUH training sets (5%, 20%~100%):

1. Fully transferred BiLSTM-CRF Hybrid: Hybrid with the BiLSTM-CRF model whose parameters of all layers were transferred from KMUH and fine-tuned on CMUH.
2. FC1-transferred BiLSTM-CRF Hybrid: Hybrid with the BiLSTM-CRF model whose parameters of the FC1 layer were transferred from KMUH and fine-tuned on CMUH.
3. Transferred HAN: The parameters of the word level and sentence level attentions were transferred from KMUH and fine-tuned on CMUH.
4. Transferred RoBERTa: The fine-tuned KMUH-RoBERTa was further fine-tuned on the CMUH corpus.

The sub-figure (b) of Fig. 4 compares the results of the above different configurations.

For the proposed hybrid systems, we noticed that the two transferred configurations outperformed their

non-transferred counterparts. Identical to the observation in the recognition task (the sub-graph (a) of Fig. 4), with only 20% of the CMUH training set, the hybrid systems with the transferred BiLSTM-CRF models achieved a promising overall F-score of 0.903, which matches with the non-transferred hybrid version trained with the full CMUH training set (*i.e.* the BiLSTM-CRF-S configuration in Table 2).

In accordance with the results shown in the sub-figure (a), the non-transferable CRF hybrid is better than that of the non-transferred BiLSTM-CRF hybrid in cases of fewer training examples. Nevertheless, the performance improvement for the CRF hybrid is limited compared to the BiLSTM-CRF hybrids when more training samples are used. The fully transferred BiLSTM-CRF hybrid demonstrated the best overall F-score of 0.911 and the performance of both the fully and FC1-transferred hybrids is even better than the best performing system displayed in Table 2. These results provide a promising insight that we can exploit the previous learned parameters of the neural networks from source hospitals to rapidly develop a reliable coding system for the new hospital by just compiling a small-scale annotated dataset containing about 60 reports. On the other hand, without transfer learning, the similar coding performance can only be achieved by leveraging the recognition models fully trained on the target hospital. The above results also validate the robustness of the established expert system for cancer registry coding when the same rule sets were applied for all configurations.

For the classification-based methods, it is intriguing to see that the two models have different behaviors. When trained with the 5% training set without transfer learning, the overall F-scores of the non-transferred HAN and BiLSTM-CRF hybrid are around 0.1. Nevertheless, the two RoBERTa models exhibited F-scores of ~ 0.4 which again substantiates the empirical strength of pre-training in transferring their general-purpose knowledge for downstream tasks within small training data [33]. We observed that by using only the 5% CMUH training set, both RoBERTa models can learn to code the H and G items with F-scores of 0.780. The non-transferred and transferred HANs can achieve similar F-scores for H and G when 80% and 40% training sets were used, respectively.

Although Ko and Choi [28] showed that multi-task fine-tuning on similar tasks is experimentally powerful for boosting the performance, our results of the two RoBERTa models do not suggest this strategy is significantly helpful for cross-hospital settings as the RoBERTa models already learned generalized universal language representations. As a result, the first fine-tuning on a dataset written with different formatting and style may not be necessary. By contrast, for the HAN model in which RoBERTa was used as a feature extractor for generating contextual representation, the pre-training on the KCMUH corpus does improve its ability to learn from the CMUH corpus. When transfer learning is employed, the HAN model outperformed the other classification-based methods including RoBERTa even with

only the 5% training data. The transferred HAN model trained with the 20% training set had a better performance than the non-transferred BiLSTM-CRF hybrid. However, when more training samples were used, the performance improvement of the hybrid systems is significantly higher than that of the HAN models. Even the non-transferred CRF hybrid system prevails over the transferred HAN, let alone the transferred hybrids. These observations altogether demonstrate the advantages of the proposed hybrid neural symbolic system.

D. ADVANTAGE AND LIMITATIONS OF TRANSFER LEARNING FOR CANCER REGISTRY ITEM RECOGNITION AND CODING

Developing a reliable AI-aided cancer registry coding system requires a large training dataset with high quality. Our experiment results illustrated in Fig. 4 reveal that with only a few training samples, the F-scores of all non-transferred systems are obviously lower than that of their full-trained counterparts. The calculation of the Pearson correlation coefficient (PCC) between the size of the training set and the overall F-scores is displayed in Table 3 which illustrates a strong correlation for all developed models excepted the transferred BiLSTM-CRF hybrid. As illustrated in Table 3, the correlation between the overall F-score of the BiLSTM-CRF-S system and the size of the training data is downgraded from strong to moderate (denoted as $S > M$ in the table) after applying transfer learning. The change of correlation coefficients after transfer learning is applied for each method and each registry item is summarized in Table 3. The interpretation of PCC is translated to descriptions based on the first table of [34]. Note that since transfer learning is not applied on CRF-S, the table only displays the PCC of each cancer registry item for the configuration of non-transferable CRF hybrid.

Our experiment results shown in the previous section have demonstrated that transfer learning enables us to reuse the knowledge learned in a source hospital to improve the performance of the coding task in a new hospital. With transfer learning, the correlation of the overall F-scores and training data sizes was reduced from strong to moderate for the developed BiLSTM-CRF-S. Although the overall performance for the other transferred models remain strongly correlated with data sizes, it can be observed that transfer learning had divergent impacts on the eight registry items for the applied methods. For the proposed hybrid neural symbolic method, transfer learning is more effective for items like TS and PN, but not for G because of the trend of the correlation change.

Unlike the proposed hybrid approaches which can benefit from transfer learning in items mainly described in numeric formats, the results in Table 3 suggest that for HAN, a classification-based approach, it is preferred to transfer knowledge learned from a source hospital for narrative items such as T, N, M, and G. However, for pre-trained language models like RoBERTa, the additional fine-tuning approach is not recommended as the analyses of PCC reveals an increase in the correlations.

TABLE 3. Pearson correlation coefficient analysis for the developed methods without and with transfer learning.

Item	BiLSTM-CRF-S ^a	CRF-S ^a	HAN ^a	RoBERTa ^a
T	M	S	V>S	S
N	M	S	M>W	S>V
M	M	S	M>N	M>S
TS	S>M	W	M	M>V
EN	S	M	S>V	M>S
PN	S>W	W	S	M>S
H	M	S	M	S>N
G	M>S	M	S>M	M>W
Overall	S>M	S	S	S

^aN = negligible correlation, W = weak correlation, M = moderate correlation; S = strong correlation, V = very strong correlation.

Taking into account the results in Fig. 4 and Table 3, we can conclude that transfer learning can relieve the requirement of obtaining a large labelled clinical corpus and the burden of the time-consuming and labor-intensive annotation process. In comparison to the classification-based models, it seems more efficient to transfer the knowledge learned from the source hospital to the target hospital through the concept recognition task and applying the proposed symbolic expert system to address the task of cancer registry coding.

1) CROSS-DATASET ANALYSIS

Although the above analysis has demonstrated that transferring parameters learned on a large concept recognition dataset for fine-tuning on a smaller one and hybridized with an expert system is likely to be a valuable solution for developing cross-hospital coding systems, it is curious to know its limitation and other possible solutions. We hence conducted a cross-dataset experiment to estimate the cross-hospital generalization ability of the developed hybrid neural symbolic systems by using the dataset from one hospital for training, and the dataset from another for testing (refer to the first column under each hospital shown in Table 4). In addition, we united the training sets from both hospitals to train a concept recognition model that can learn the characteristics of the reports from both hospitals simultaneously when compared to the transferred model and hybridized it with the same symbolic expert system (the KMHU + CMUH-S columns shown in Table 4). The performance of the hybrid with the transferred KMHU model on CMUH and its source hospital (KMHU) is listed in the second columns denoted as KMHU-T-S.

Although both the recognition corpora were annotated by the same annotators following the same annotation guideline, we noticed that the generality of the developed BiLSTM-CRF models for the recognition task is not well as the overall F-scores dropped to 0.541 and 0.459 on the test sets of KMHU and CMUH, respectively. By hybridized the same expert system, the coding performance is still outperformed the classification-based approaches shown in Table 2, but a significantly drop in F-score can be observed. Overall, only two registry items (EN and PN) achieved acceptable recognition F-scores (0.792 to 0.921) in the cross-hospital setting, while the F-scores of all others are lower than 0.58.

These results uncover the fact that the format and writing styles of the descriptive pathology in surgical biopsy reports across hospitals are heterogeneous in real-world scenarios. Comparing with the token-level evaluation for the recognition task, the document-level coding performance is much better, which demonstrates that 1) the extracted registry concept information can be effectively exploited by the developed symbolic coding system in the cross-hospital setting; 2) given that the coding results were individually curated by registrars in the two hospitals following the TCR coding manual, the consistent and robust results of the developed expert system highlight the high quality of curated results.

Finally, we found that the model trained with the merged dataset achieved the best recognition F-scores on both test sets, which even outperformed models trained using only the dataset from the same target hospital. Although the transferred model obtained an F-score of 0.954 on the CMUH test set, which is tied with the merged model, the F-score on its source test set KMHU was decreased by 0.18 as all registry items except EN and PN have decreased F-scores. In particular, the performance of the G item dropped significantly from 1.00 to 0.516. The results match with the results of our PCC analysis shown in Table 3. This finding demonstrates that the transferred recognition model suffered the catastrophic forgetting problem [35] which is known to be a challenge for artificial neural networks when the network is trained sequentially on multiple tasks because the weights in the network that are important for the original hospital are now changed to meet the objectives of the data from the new hospital. In contrast, the impact of the catastrophic forgetting for the hybridized system is much less significant. By comparing with Table 2, we can see that the coding performance is the same or even better for N, H and EN items, which demonstrates the robustness of the proposed hybrid neural symbolic approach.

E. ERROR ANALYSIS

As shown in Fig. 2, our implementation of the expert system supports goal-driven reasoning to provide an explanation facility for the coding results. During the development of our hybrid neural symbolic system, we used the facility to examine the training/development sets and refine the rules established based on the interpretations of the TCR coding manual. Although TCR is known to be one of the high quality cancer registries worldwide [29], we still identified some coding or content errors in our corpus during system development. These errors were mainly caused by human factors including coding errors by registrars (*e.g.* the given codes were inconsistent with the contents of all available reports) and inconsistent reporting by pathologists (*e.g.* narrative descriptions of cancer grades varied across the reports of a patient in the course of the treatment). Those inconsistencies increase the difficulty of the developed expert systems for processing multiple reports associated with one patient. Both cases result in the increase of both FP and FN for coding by 1. Fortunately, the explanation facility of our expert system

TABLE 4. Cross-corpus evaluation of the proposed hybrid neural symbolic systems for the task of concept recognition and coding.

Hospital	CMUH						KMUH					
Dataset	KMUH-S ^a		KMUH-T-S ^b		KMUH+CMUH-S ^c		CMUH-S ^d		KMUH-T-S ^b		KMUH+CMUH-S ^c	
Task	R ^e	C ^f	R ^e	C ^f	R ^e	C ^f	R ^e	C ^f	R ^e	C ^f	R ^e	C ^f
T	0.502	0.644	0.983	0.939	0.996	0.939	0.517	0.860	0.835	0.881	0.980	0.883
N	0.258	0.200	0.959	0.960	0.967	0.960	0.370	0.842	0.852	0.912	0.984	0.912
M	0.000	0.800	0.971	0.853	0.967	0.853	n/a	0.982	n/a	0.982	n/a	0.982
TS	0.156	0.345	0.932	0.870	0.949	0.875	0.506	0.632	0.748	0.807	0.969	0.842
EN	0.792	0.909	0.944	0.910	0.946	0.910	0.800	0.684	0.987	0.860	0.987	0.860
PN	0.858	0.933	0.963	0.960	0.966	0.960	0.921	0.877	0.990	0.877	0.993	0.877
SC	0.500	n/a	0.972	n/a	0.978	n/a	0.455	n/a	0.856	n/a	0.990	n/a
H	0.466	0.839	0.919	0.902	0.905	0.894	0.406	0.916	0.857	0.956	0.987	0.956
G	0.268	0.400	0.940	0.885	0.942	0.881	0.000	0.519	0.516	0.881	1.000	0.930
Overall	0.459	0.615	0.954	0.911	0.954	0.903	0.541	0.756	0.804	0.897	0.986	0.910

^aKMUH-S = the model trained on the KMUH training set hybridized with the symbolic expert system.

^bKMUH-T-S = the KMUH model fine-tuned on the CMUH training set hybridized with the symbolic expert system.

^cKMUH+CMUH-S = the model trained on the union of the training sets of KMUH and CMUH hybridized with the symbolic expert system.

^dCMUH-S = the model trained on the CMUH training set hybridized with the symbolic expert system.

^eR = recognition task (subtask 1).

^fC = coding task (subtask 2).

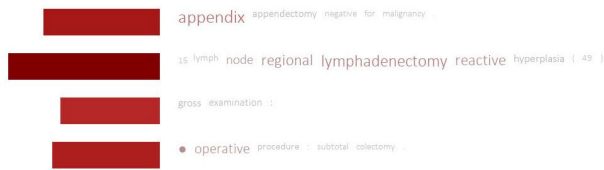


FIGURE 5. The visualization of the attention weights in the HAN model for PN coding. The size of a word represents the weight for making the prediction. In this example, the model can successfully generate a correct code of 0 based on the excerpted narrative description.

provides us and registrars a tool to uncover the inconsistency and help to diagnose the curated results to improve the quality of curated data by highlighting unknown curation errors.

Moreover, we identified three main categories of errors for the proposed hybrid neural symbolic system. In the first category, the coding error is due to the recognition errors of the underline neural network. The second category owes to incomplete or insufficient information associated with the target registry code in the patient’s reports, thus it is not possible for the proposed system or other baseline models to correctly generate the target codes. For example, one of the coding rules for TS is that if the patient received neoadjuvant therapy, the code should be recorded based on the original size of the tumor before the therapy. However, the information of the original size is not always available in the reports.

The last category is due to incomplete rule definitions that can be refined in the future by introducing more cancer registry-related concepts which can be extracted as facts for complementing the current knowledge representation. For instance, the current rules defined for PN did not check whether the described numbers of nodes are benign or malignant. Therefore, an incorrect code of 49 will be suggested for the following description: “15. Lymph node regional lymphadenectomy reactive hyperplasia (49)”. On the other hand, we noticed that the HAN model seems to learn to capture the key concept “reactive hyperplasia” to address the aforementioned issue of the developed hybrid system as shown in Fig. 5. However, the HAN model failed in cases where

all individually described numbers of examined/positive nodes need to be summed up to generate the corresponding EN/PN codes.

IV. CONCLUSION

A national cancer registry is essential for monitoring the cancer incidence and mortality and for developing strategies and policies to control the disease. Both completeness and accuracy are important quality measures for the registry database. We have demonstrated the feasibility of the hybrid of neural and symbolic AI for the task of cancer registry coding and the robustness of the developed system. We confirmed that the different writing styles and formats of the pathology reports among hospitals indeed diminish the performance of the developed systems developed from the dataset of one hospital to another. Factors that influence the performance depend on the characteristics of the registry items and the underlying methods. Overall, when transfer learning is adopted, the model pre-trained on a source hospital can be trained with fewer annotations of the target hospital and achieve similar performance when the full training set of the target hospital is used without transfer learning. Our empirical results indicated that the proposed hybrid approaches mainly benefitted from transfer learning in items described in numeric formats, whereas it is preferred to transfer knowledge learned for narrative items for classification-based approaches like HAN. For pre-trained language models, our experiment results imply that additional fine-tuning is not necessary. Moving forward, we will seek to collaborate with more hospitals in Taiwan to extend the knowledge of the developed system for diverse cancers.

REFERENCES

- [1] L. X. Clegg, “Impact of reporting delay and reporting error on cancer incidence rates and trends,” *CancerSpectrum Knowl. Environ.*, vol. 94, no. 20, pp. 1537–1545, Oct. 2002.
- [2] S. Gao, M. T. Young, J. X. Qiu, H.-J. Yoon, J. B. Christian, P. A. Fearn, G. D. Tourassi, and A. Ramanathan, “Hierarchical attention networks for information extraction from cancer pathology reports,” *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 3, pp. 321–330, Mar. 2018.

- [3] M. Alawad, S. Gao, J. X. Qiu, H. J. Yoon, J. Blair Christian, L. Penberthy, B. Mumphy, X.-C. Wu, L. Coyle, and G. Tourassi, "Automatic extraction of cancer registry reportable information from free-text pathology reports using multitask convolutional neural networks," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 1, pp. 89–98, Jan. 2020.
- [4] K. L. Kehl, H. Elmarakeby, M. Nishino, E. M. Van Allen, E. M. Lepisto, M. J. Hassett, B. E. Johnson, and D. Schrag, "Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports," *JAMA Oncol.*, vol. 5, no. 10, pp. 1421–1429, 2019.
- [5] J. X. Qiu, H.-J. Yoon, P. A. Fearn, and G. D. Tourassi, "Deep learning for automated extraction of primary sites from cancer pathology reports," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 244–251, Jan. 2018.
- [6] K. Roberts, Y. Si, A. Gandhi, and E. Bernstam, "A framework for cancer information in clinical narratives: Schema and annotation," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, May 2018, pp. 1–8.
- [7] G. K. Savova, E. Tseytlin, S. Finan, M. Castine, T. Miller, O. Medvedeva, D. Harris, H. Hochheiser, C. Lin, G. Chavan, and R. S. Jacobson, "DeepPhe: A natural language processing system for extracting cancer phenotypes from clinical records," *Cancer Res.*, vol. 77, no. 21, pp. e115–e118, Nov. 2017.
- [8] A. Coden, G. Savova, I. Sominsky, M. Tanenblatt, J. Masanz, K. Schuler, J. Cooper, W. Guan, and P. C. de Groen, "Automatically extracting cancer disease characteristics from pathology reports into a disease knowledge representation model," *J. Biomed. Informat.*, vol. 42, no. 5, pp. 937–949, Oct. 2009.
- [9] S. B. Edge and S. B. Edge, *AJCC Cancer Staging Manual*, 8th ed. New York, NY, USA: Springer, 2017.
- [10] Y.-J. Lin, H.-J. Dai, Y.-C. Zhang, C.-Y. Wu, Y.-C. Chang, P.-J. Lu, C.-J. Huang, Y.-T. Wang, H.-M. Hsieh, K.-S. Chao, T.-W. Liu, I.-S. Chang, Y.-H.-C. Yang, T.-H. Wang, K.-J. Liu, L.-T. Chen, and S.-F. Yang, "Cancer registry information extraction via transfer learning," in *Proc. 3rd Clin. Natural Lang. Process. Workshop*, 2020, pp. 201–208.
- [11] S. Wermter and R. Sun, "An overview of hybrid neural systems," in *Proc. Int. Workshop Hybrid Neural Syst.* Berlin, Germany: Springer, 1998, pp. 1–13.
- [12] H.-J. Dai, S. Syed-Abdul, C.-W. Chen, and C.-C. Wu, "Recognition and evaluation of clinical section headings in clinical documents using token-based formulation with conditional random fields," *BioMed Res. Int.*, vol. 2015, Aug. 2015, Art. no. 873012.
- [13] X. Zhang, Y. Zhang, Q. Zhang, Y. Ren, T. Qiu, J. Ma, and Q. Sun, "Extracting comprehensive clinical information for breast cancer using deep learning methods," *Int. J. Med. Informat.*, vol. 132, Dec. 2019, Art. no. 103985.
- [14] T. Oliwa, S. B. Maron, L. M. Chase, S. Lomnicki, D. V. T. Catenacci, B. Furner, and S. L. Volchenbom, "Obtaining knowledge in pathology reports through a natural language processing approach with classification, named-entity recognition, and relation-extraction heuristics," *JCO Clin. Cancer Informat.*, vol. 3, pp. 1–8, Dec. 2019.
- [15] T. Mostafiz and K. Ashraf, "Pathology extraction from chest X-ray radiology reports: A performance study," 2018, *arXiv:1812.02305*. [Online]. Available: <http://arxiv.org/abs/1812.02305>
- [16] X. Han and R. Ruonan, "The method of medical named entity recognition based on semantic model and improved SVM-KNN algorithm," in *Proc. 7th Int. Conf. Semantics, Knowl. Grids*, Oct. 2011, pp. 21–27.
- [17] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
- [18] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019, *arXiv:1907.11692*. [Online]. Available: <https://arxiv.org/abs/1907.11692>
- [19] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, "Hierarchical attention networks for document classification," in *Proc. NAACL*, 2016, pp. 1480–1489.
- [20] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Int. J. Data Warehousing Mining*, vol. 3, no. 3, pp. 1–13, 2007.
- [21] H.-J. Dai, C.-H. Su, Y.-Q. Lee, Y.-C. Zhang, C.-K. Wang, C.-J. Kuo, and C.-S. Wu, "Deep learning-based natural language processing for screening psychiatric patients," *Frontiers Psychiatry*, vol. 11, p. 1557, Jan. 2021.
- [22] D. Martinez and Y. Li, "Information extraction from pathology reports in a hospital setting," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2011, pp. 1877–1882.
- [23] J. Turian, L. Ratinov, and Y. Bengio, "Word representations: A simple and general method for semi-supervised learning," in *Proc. 48th Annu. Meeting Assoc. Comput. Linguistics*, 2010, pp. 384–394.
- [24] H.-J. Dai, P.-T. Lai, Y.-C. Chang, and R. T.-H. Tsai, "Enhancing of chemical compound and drug name recognition using representative tag scheme and fine-grained tokenization," *J. Cheminformatics*, vol. 7, no. S1, p. S14, Jan. 2015, doi: [10.1186/1758-2946-7-s1-s14](https://doi.org/10.1186/1758-2946-7-s1-s14).
- [25] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," presented at the NAACL-HLT, Minneapolis, MN, USA, 2019.
- [26] A. Miranda-Escalada, E. Farré, and M. Krallinger, "Named entity recognition, concept normalization and clinical coding: Overview of the cantemist track for cancer text mining in Spanish, corpus, guidelines, methods and results," in *Proc. Iberian Lang. Eval. Forum (IberLEF), CEUR Workshop*, Sep. 2020, pp. 303–323.
- [27] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [28] B. Ko and H. Choi, "Twice fine-tuning deep neural networks for paraphrase identification," *Electron. Lett.*, vol. 56, no. 9, pp. 444–447, Apr. 2020.
- [29] C.-J. Chiang, S.-L. You, C.-J. Chen, Y.-W. Yang, W.-C. Lo, and M.-S. Lai, "Quality assessment and improvement of nationwide cancer registration system in taiwan: A review," *Jpn. J. Clin. Oncol.*, vol. 45, no. 3, pp. 291–296, Mar. 2015.
- [30] E. Wallace, Y. Wang, S. Li, S. Singh, and M. Gardner, "Do NLP models know numbers? Probing numeracy in embeddings," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 5310–5318.
- [31] J. M. Giorgi and G. D. Bader, "Transfer learning for biomedical named entity recognition with neural networks," *Bioinformatics*, vol. 34, no. 23, pp. 4087–4094, Dec. 2018.
- [32] J. M. Giorgi and G. D. Bader, "Towards reliable named entity recognition in the biomedical domain," *Bioinformatics*, vol. 36, no. 1, pp. 280–286, Jan. 2020.
- [33] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification," in *China Nat. Conf. Chin. Comput. Linguistics*. Cham, Switzerland: Springer, 2019, pp. 194–206.
- [34] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia Analgesia*, vol. 126, no. 5, pp. 1763–1768, May 2018.
- [35] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends Cogn. Sci.*, vol. 3, no. 4, pp. 128–135, Apr. 1999.



HONG-JIE DAI (Member, IEEE) received a Ph.D. degree in computer science from the National Tsing Hua University in 2012. He became a Member of IEEE in 2017.

From 2013 to 2015, he was an Assistant Professor at the Graduate Institute of Biomedical Informatics in Taipei Medical University (TMU). He served as the director of the Application Development Section of National Taitung University (NTTU) Library and Information Centre and an Associate Professor with the Department of Computer Science and Engineering, NTTU from 2016 to 2017. He was a Visiting Scholar at the Institute of Information Science, Academia Sinica, Taiwan from 2013 to 2018. Since 2019, he has been an Adjunct Associated Research Fellow with National Institute of Cancer Research, National Health Research Insurance, Taiwan and a Jointly Appointed Associate Professor of the Kaohsiung Medical University, Taiwan. He is currently a Professor with the Department of Electrical Engineering in National Kaohsiung University of Science and Technology (NKUST), Taiwan. He has authored book chapters for biomedical and clinical text mining and has published research papers in peer-reviewed journals on fields including natural language processing, bioinformatics, and clinical informatics. His current research interests include natural language processing, artificial intelligence, bioinformatics, medical informatics, machine learning, and software engineering.

Dr. Dai is a member of the Taiwan Association for Medical Informatics, Taiwanese Association for Artificial Intelligence, and the Association for Computational Linguistics and Chinese Language Processing. He was a recipient of the Best Doctoral Dissertation Award for the Association for Computational Linguistics and Chinese Language Processing in 2012, TMU Newly-Appointed National Science Council Research Fellow Award in 2013 and 2014, NTTU Outstanding Research Merit Award in 2016 to 2018 and NKUST Outstanding Research Merit Award in 2020.



YI-HSIN YANG currently is an Investigator in the National Institute of Cancer Research, National Health Research Institutes, and an adjunct Professor in School of Pharmacy, Kaohsiung Medical University, Taiwan. She received her PhD degree in Biostatistics from University of North Carolina at Chapel Hill, USA. Professor Yang has been devoted to promote research on real-world data (RWD) and real-world evidence (RWE) through the use of health administrative database and government database (particularly, the National Health Insurance Research Database, Cancer Registry Database, Cancer Screening Database, National Health Interview Survey, etc.). Her focus includes analysis methodology and clinical research using health database. While maximize data impact has become an important theme of current cancer control program. She has been actively involved in the field of pharmacoepidemiology. She published several articles in chemoprevention agents for cancers as well as comparative effectiveness on cancer treatment modalities.



TI-HAO WANG received his Doctor of Medicine degree from National Taiwan University in 2010. He has been an attending physician of the Division of Radiation Oncology, Department of Oncology, Taipei Veterans General Hospital, Taipei, Taiwan. Ti-Hao Wang is on a mission to bridge state-of-art informatic technology to medicine. Trained as a radiation oncologist, he has a solid background in cancer care. What makes him unique is during his clinical career, he developed several integrated workflow systems to help clinicians increase efficiency. He also has experience in deep learning, computer vision, and natural language processing. Currently, he works in China Medical University Hospital as an attending physician and Everfortune.AI company as R&D director.



YAN-JIE LIN received his Ph.D. degree in physiology from the School of Medicine, National Yang-Ming University (NYMU, now National Yang Ming Chiao Tung University, NYCU) in 2014. Dr. Lin was a recipient of the Best Thesis Award of Taiwan Society of Lipid and Atherosclerosis in 2015. He is a member of Federation of the Asian & Oceanian Physiological Societies, and focus on experimental and clinical medicine. Since 2015, Dr. Lin has become a Postdoctoral Research Fellow in the National Institute of Cancer Research (NICTR), and then in the Institute of Population Health Sciences (IPHS), National Health Research Institutes (NHRI), Taiwan.



PIN-JOU LU is currently studying at the Electrical Engineering Department of National Kaohsiung University of Science and Technology, Taiwan. She served as a knowledge engineer of the expert system developed for the Integrated Cancer Research Multicenter Collaboration Platform project.



CHUNG-YANG WU is currently studying at the Electrical Engineering Department of National Kaohsiung University of Science and Technology, Taiwan. He served as an artificial intelligence developer of the Integrated Cancer Research Multicenter Collaboration Platform project.



YU-CHENG CHANG is currently studying at the Electrical Engineering Department of National Kaohsiung University of Science and Technology, Taiwan. He served as a user interface developer of the Integrated Cancer Research Multicenter Collaboration Platform project.



YOU-QIAN LEE was born on July 27th, 1998 in New Taipei city, Taiwan. He received his B.S. in Engineering in the Department of Electrical Engineering from the National Kaohsiung University of Science and Technology (NKUST), Taiwan. He is currently working on his Master degree in Electrical Engineering at NKUST. His research interests include electronic health records, code-switch analysis and de-identification.



YOU-CHEN ZHANG received his B.S. degree in Computer Science and Information Engineering from National Taitung University, Taiwan. In 2020, he served as a natural language processing developer of the Integrated Cancer Research Multicenter Collaboration Platform project. His research interests lie in the area of natural language processing and deep learning. He is studying for his master's degree in Electrical Engineering at National Kaohsiung University of Science and Technology, Taiwan/ His current research focuses on solving the catastrophic forgetting problem.



YUAN-CHI HSU is currently studying in the Department of Electrical Engineering at National Kaohsiung University of Science and Technology, Taiwan. He served as an expert system developer of the Integrated Cancer Research Multicenter Collaboration Platform project. He is also working on the development of a natural language processing-aided cancer registry database and web interface for a related project.



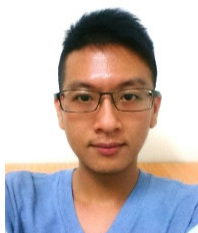
HAN-HSIANG WU is currently studying at the Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Taiwan. He served as an annotator of the Integrated Cancer Research Multicenter Collaboration Platform project. His current research interests include data analysis and natural language processing.



CHENG-RONG KE is studying in the Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Taiwan. He served as a natural language processing (NLP) developer of the Integrated Cancer Research Multicenter Collaboration Platform project leading by the National Institute of Health. He is working on the development of an NLP pipeline for automatically analyzing of medical records, and the web application programming interface.



CHI-HJEN HUANG received M.D. degree and the Ph.D. degree in Medicine from Kaohsiung Medical University, Kaohsiung, Taiwan, ROC, in 1989 and 2012, respectively. He is currently a Professor with Radiation Oncology in Kaohsiung Medical University. His research interests include tumor hypoxia, radiotherapy specialized in breast cancer and head neck cancer. He also served as the Chair of department of Radiation Oncology and Cancer Center in Kaohsiung Medical University Hospital.



YU-TSANG WANG was born in Kaohsiung, Taiwan in 1987. He received a M.S. degree in Public Health from the China Medical University in 2011. From 2011 to 2014, He has worked in the Infection Control and Biosafety in the Taiwan Centers for Disease Control, Taiwan, where he was an Research Assistant. Starting from 2014, he is a Statistical Analyst in the Division of Medical Statistics and Bioinformatics, Kaohsiung Medical University Chung-Ho Memorial Hospital, Taiwan.

His current research interests include Public Health, Epidemiology, Biostatistics and Health data analysis.



SHEAU-FANG YANG received the B.S. degree in medicine from China Medical College, Taichung, Taiwan, in 1996 and the M.S. degree in Science from Kaohsiung Medical University, Kaohsiung, Taiwan, in 2004. She is currently an Associate Professor of the School of Medicine, Kaohsiung Medical University, and also an Attending Physician in the Department of Pathology, Kaohsiung Medical University Chung-Ho Memorial Hospital, Taiwan. Her research interests include surgical pathology, breast oncology and hepatic oncology.



KUAN-CHUNG HSIAO received the B.S degree in medical technology in 2002 and the M.S degree in microbiology and immunology in 2004 from National Cheng Kung University, Tainan, Taiwan. He received the Ph.D. degree in tumor immunology from National Cheng Kung University, Tainan, Taiwan in 2013.

He was a Research Assistant with the National Institute of Cancer Research, National Health Research Institutes, Taiwan from 2005 to 2009.

Since 2013, he has been a Postdoctoral Fellow with the National Institute of Cancer Research, National Health Research Institutes, Taiwan. His research interests include tumor immunology and cancer immunotherapy, mechanisms of tumor metastasis and the imbalance of osteogenesis.



KO-JIUNN LIU received the B.S degree in medical technology from Taipei Medical College, Taipei, Taiwan in 1986 and the Ph.D. degree in tumor immunology from Northwestern University Evanston, IL in 1994.

From 1994 to 1997, he was a Postdoctoral Fellow with the Division of Oncology, University of Washington, USA. He has been an Assistant Investigator with the National Institute of Cancer Research, National Health Research Institutes, Taiwan in 1997 and he has been the Deputy Director with the National Institute of Cancer Research, Taiwan since 2018. His research interests include tumor immunology and cancer immunotherapy, identification of tumor markers and genetic factors in cancers.

Dr. Liu was a recipient of the Taipei Biotech Award in 2018.



LI-TZONG CHEN received his M.D. and Ph.D. degree in Kaohsiung Medical University (KMU), Kaohsiung, Taiwan. Dr. Chen is a medical oncologist and gastroenterologist, and joined National Institute of Cancer Research (NICR), National Health Research Institutes (NHRI), in 1995. As a leading clinical researcher in NICR cancer ward, he has established NICR clinical research team as one of the most creditable early phase clinical research teams in Asia. Dr. Chen is the Distinguished Investigator and the Director of NICR, NHRI, and co-affiliated as the Chair Professor of Internal Medicine, Kaohsiung Medical University Hospital, KMU, Taiwan.

Dr. Chen's research interests focus on translational research and experimental therapy to develop new management strategies for gastrointestinal (GI) cancers. He has participated in the development of several upper GI cancer management guidelines in Asia. Dr. Chen has organized inter-institution translational research teams under Taiwan Cooperative Oncology Group (TCOG) working on hepatocellular, pancreatic, biliary tract cancer and rare GI tumors, such as gastrointestinal stromal tumor (GIST) and neuroendocrine tumor (NET) through various investigator-initiated, multicenter cooperative studies, and has been involved in multiple industry-sponsored clinical trials worldwide.



I-SHOU CHANG is an Emeritus Investigator of the National Institute of Cancer Research, National Health Research Institutes, Taiwan. His current research interests include genome-wide association studies, risk prediction models, epidemiology of cancer, and related statistics methodologies.



K. S. CLIFFORD CHAO is a pre-eminent expert in the use of image-guided targeted radiotherapy for the treatment of cancers and has served as Principal Investigator of multiple NIH and DoD research grants and published more than 190 peer-reviewed papers. He was a tenured associate professor at Washington University School of Medicine in St. Louis, a professor at University of Texas, MDACC, and later became distinguished chaired Professor and the Department Chair of

Radiation Oncology of Columbia University and Weill Cornell Medical College. Dr. Chao's research focus includes combining positron emission tomography (PET) and computed tomography (CT) images to construct individualized treatment plans that address the molecular characteristics of each tumor. He co-founded two MD Anderson-based start-up company using a knowledge-based artificial intelligence to guide high precision robotic image-guided therapy system that helps map the exact location of a tumor and direct radiation towards the malignant cells while avoiding healthy tissue. From 2014-Present, he serves a Superintendent of the Cancer Center, China Medical University Hospital, Taiwan. His research includes the development of translational research on the immunoregulation for the treatment of colorectal cancer, breast cancer, lung cancer and esophageal cancer. His laboratory is working on discovery of shared tumor neoantigens in tumor that failed standard treatment. He participates in the projects of multi-center clinical information database, AI-based colon cancer research cloud platform. He devotes in the field of tumor neoantigen and AI medical artificial intelligence.

Dr. Chao, from 2016 to 2020, served as the President Taiwan Society for Therapeutic Radiology and Oncology. He has also received numerous academic as well as Best Doctors in America awards.



TSANG-WU LIU received his M.D. degree from the National Taiwan University, Taiwan, in 1981. He completed the medical oncology/Hematology fellowship training in July 1989 from the Department of Internal Medicine, National Taiwan University Hospital and Academic Sinica's Medical Oncology Fellowship Program (program director: late Professor Paul Carbone). Dr. Liu is the investigator and currently serves as the Deputy Director of the NICR, NHRI, Taiwan.

Dr. Liu became deputy chairman of Advisory Committee (AC) in 2004, leading the Investigational New Drug (IND) from 2005 to 2008, and then he became the chairman of the Advisory Committee to Bureau of Pharmaceutical Affairs (BPA). Invited by Bureau of Health Promotion (BHP, now Health Promotion Administration), Dr. Jacqueline Whang-Peng and Dr. Liu develop the Accreditation System for cancer care program in Taiwan since 2005, and Dr. Liu is the principal investigator of this project since 2008. Another project Dr. Liu have spent effort is the Project of Developing and Monitoring Core Measurement Indicators for Cancer Screening and Treatment which utilizes the Taiwan Cancer Registry Database and the hospital report data to monitor the quality of cancer care in Taiwan, and to provide recommendations to Health Promotion Administration.

• • •