

# 深層学習を用いた電子カルテ医療情報の多角的解析

衛藤 亮太<sup>†1,a)</sup> 松原 靖子<sup>†1</sup> 山下 和人<sup>†2</sup> 國澤 進<sup>†2</sup> 今中 雄一<sup>†2</sup> 櫻井 保志<sup>†1</sup>

**概要:** 本研究では、電子カルテ医療情報を多角的に解析するための深層学習を用いた医療情報分類モデルについて述べる。提案手法は、利用者が受診した医療機関や診断された疾患および手術方法などの多次元情報を含む電子カルテ医療情報が与えられたときに、その医療情報から分類に有用な特徴を発見し、利用者の予後を表す転帰情報の分類を行う。実データを用いた実験では、提案手法が多次元医療情報の中から分類に有用な情報を共起影響を加味した上で特定することを確認し、既存の分類モデルとの比較を行い提案手法の精度が向上をしていることを示した。

**キーワード:** 深層学習, 特徴選択, 分類, 医療情報

## Analysis of electronic health record using Deep Learning

RYOTA ETO<sup>†1,a)</sup> YASUKO MATSUBARA<sup>†1</sup> KAZUTO YAMASHITA<sup>†2</sup> SUSUMU KUNISAWA<sup>†2</sup>  
YUICHI IMANAKA<sup>†2</sup> YASUSHI SAKURAI<sup>†1</sup>

**Abstract:** In this paper, we describe a medical information classification model using deep learning to analysis electronic health record(EHR). In the proposed method, when EHR including multi dimensional information such as users's disease and surgery method is given, from the medical information, special medical information necessary for the classification are found and the outcome information of the user is classified. In experiments using EHR data, we confirmed that the proposed method specifies useful information for classification from multi dimensional medical information, compared with existing classification model, and showed high accuracy of the proposed method

**Keywords:** DeepLearning, Feature Selection, Classification, Medical information

### 1. はじめに

近年、先進国を中心として高齢者人口の増加や出生率の低下により少子高齢化の進行、ならびに国民医療費の増大といった数多くの問題が社会保障制度に大きく影響を与えている。特に我が国は、先進国の中で高齢化率が最も高く、持続可能な社会保障制度の構築のための対策が急務と

されている。これらの対策の一つとして挙げられているのが、昨今の情報化社会の発展に伴い生じた診療報酬明細書やCT、MRIなどの診断機器から取得した様々な大規模医療情報の活用である。また医療情報を用いた研究の中でも、2012年にILSVRC<sup>\*1</sup>で発表された従来手法を大きく上回る画像認識精度を得た深層学習(Deep Learning)を利用した研究事例が特に注目されている。深層学習は機械学習アルゴリズムの一つであり、多層ニューラルネットワーク(Deep Neural Network)[1]を学習することでデータの hoch 表現を獲得することが可能なことで知られており、画像認識や音声認識の分野で数多くの研究事例が報告されている。しかし、医療情報に対し深層学習を適用した研究事例はCTやMRI画像などの医用画像を解析対象とするもの

<sup>†1</sup> 現在、熊本大学 大学院自然科学研究科 情報電気電子工学専攻  
Presently with Department of Computer Science and Electrical Engineering Graduate School of Science and Technology Kumamoto University

<sup>†2</sup> 現在、京都大学 大学院医学研究科 医療経済学分野  
Presently with Department of Healthcare Economics and Quality Management Graduate School of Medicine Kyoto University

a) 161d8620@st.kumamoto-u.ac.jp

<sup>\*1</sup> <http://www.image-net.org/challenges/LSVRC/>

が非常に多い。これは医用画像を用いる場合、画像認識能力に長けた畳み込みニューラルネットワーク (Convolution Neural Networks:CNN) の適用が所望の結果を得ることが広く認知されていることが起因している。これらの情報技術は医療に対して大きな役割を果たしているが、一方で医療機関や社会全体の医療健康増進のための有力な医療情報解析手法は未だ確立されていないのが実情である。

今後は、情報技術を医療の質の向上や医療資源配分の適正化等に繋げるため、電子カルテのような複雑な医療情報を多角的に解析可能な技術が社会に重要な役割を果たすと考えられる。そこで本研究では電子カルテを多角的に解析することを目的とする。より具体的には、病名や手術などの多様な情報を持つ電子カルテ情報を扱い、患者の予後を示す転帰等の分類を行うため深層学習を用いた分類モデルを提案する。また医療情報の解析に伴い分類を行うモデルへの入力情報の選定は非常に重要な課題である。そのため本研究では、分類モデルの提案に加えて電子カルテ医療情報から分類に有用な特徴の選定を行う。本論文の構成は以下の通りである。2節では深層学習および特徴選択について述べる。3節では関連研究、4節では提案手法、5節では実データを用いた実験結果について述べ、6節にて本稿をまとめる。

## 2. 背景

### 2.1 深層学習

深層学習は多層のニューラルネットワークを用いた機械学習のアルゴリズムである。深層学習は幾つかの方法論の集まりであり、適用する問題によって方法論を使い分けるのが一般的である。画像認識の分野には視野脳の情報伝達システムをニューラルネットに適用したCNNが用いられる。次元削減や特徴抽出問題には深層自己符号器 (Deep Auto Encoder:DAE) 等が適用される。また音声や言語、動画などの系列データ処理には再帰型ニューラルネットワーク (Recurrent Neural Networks:RNN) が活用される。本研究ではRNNを利用したモデルを構成する。RNNは有向閉路を持つネットワークであり隠れ層に閉路を設置することで、過去の入力情報を現行に反映できる。そのため音声や言語といった時系列データに適用される。またRNNは有向閉路層を時系列順に展開することで、順伝搬型ネットワークとして扱うことができる。

RNNは有向閉路を持ち、各層間で異なる重みパラメータを保持する。次にある変数 $t$ でのRNNの出力を示す順伝搬計算を(1)式に示す。

$$\mathbf{z}^t = \mathbf{f}(\mathbf{W}^{In-Hi} \mathbf{x}^t + \mathbf{W} \mathbf{z}^{t-1}) \quad (1)$$

(1)式において $\mathbf{z}$ は出力、 $\mathbf{x}^t$ は $t$ におけるRNNへの入力を示す。(1)式から、RNNは過去全ての入力情報を反映

した結果が得られると考えられるが、通常のRNNは順伝搬型ネットワークの勾配消失問題によって反映可能な過去情報数に限りがある。

### 2.2 特徴選択

分類器を作成する際、どのような情報を分類器の学習に用いるかという問題はクラス間で違いを示す医療情報を見つけることに相当し分類器の性能に大きく影響する。特徴選択法は、入力情報が保有する性質を利用するFilter法と選択する特徴部分集合を変化させながら各特徴の評価を行うWrapper法に大別される。一般的にWrapper法は、実際の分類器により特徴の評価を行うため各特徴の評価である識別率は高くなるが、多量の分類器を構成するため計算量が多い。一方でFilter法は計算量は少ないが識別率は低くなるとされている。また電子カルテの多次元医療情報の分類においては、分類対象である転帰等の種別の多様性に影響を受けない特徴に加えて、複数の特徴による共起影響を考慮する必要がある。複数特徴の共起影響を考慮したFilter法に基づく特徴選択手法としてMCFS(Multi Cluster Feature Selection)[2]がある。MCFSはデータセットから分類精度に起因する特徴を選択することができる。しかし、本研究で用いる電子カルテ情報は名義変数型の情報が多数存在すると共に、多数の特徴から構成される属性情報の選択を行うことができない。

## 3. 関連研究

### 3.1 医療情報解析

医療情報に対して情報処理を適用した研究はX線やMRIなどの医用画像に加えて、心電図や脳波といった生体信号処理など多岐に渡って行われている。医療情報に対してクラス分類を試みた事例にはランダムフォレスト (Random Forests) やサポートベクターマシン (Support Vector Machine:SVM) が利用されている [3]。特にSVMはパターン認識能力に優れたモデルであり、One against reset 法等を用いることによって多クラス分類を実現できる。またカーネル関数を用いて定義される非線形変換によって非線形分類器に拡張できる高い汎化性能を持つ。次に近年、注目を浴びているDeep Learningを医療情報解析に適用した研究事例について述べる。

### 3.2 深層学習の医療情報への適用

深層学習の医療情報への適用事例は医用画像に対するものが多くの割合を占めている。しかし使用されるネットワークは画像認識処理に優れたCNNのみではない。文献 [8] では再帰型ニューラルネットワークを用いて画像認識に用いるラベルを自動的に付与する研究が報告されている。また特徴抽出モデルであるVAE(Variational Auto

表 1 転帰区分

転帰	区分
1	傷病が治癒したと判断される場合
2	傷病が軽快したと判断される場合
3	傷病(白血病等)が寛解したと判断される場合
4	傷病が不変と判断させる場合
5	傷病が悪化したと判断される場合
6	傷病による死亡
7	傷病以外による死亡

Encoder:VAE) を用いて胸部レントゲン写真から胸部異常を検出することに成功している。続いて本研究で提案する手法について述べる。まず解析対象となる電子カルテより抽出した医療情報について述べる。

## 4. 提案手法

### 4.1 医療情報

本研究では、京都府の複数病院より提供を受けた電子カルテ情報を用いる。電子カルテには、病名や術式を文字や数値で表現するために ICD10 コードや K コードが用いられている。そこで ICD10 並びに K コード等によって記録されている属性を代表して dmain, outcome, opek1 のそれぞれについて述べる。また dmain 等を電子カルテ情報の属性と呼び、本研究で扱う電子カルテには 24 種の属性が存在する。

#### (1) 主病名 (dmain)

dmain は患者が診断された主な疾病の名称が ICD10 コードによって記載されている。ICD(International Statistical Classification of Diseases and Related Health Problems) は、死因や疾病の統計基準として世界保健機構によって作成された国際統計分類の指標である。ICD10 コードは、アルファベットと数字で構成される。例えば、"Q801" という ICD10 コードは先天性および染色体が要因とされる疾病であり X 連鎖性劣性魚鱗癬という病が該当する。

#### (2) 転帰 (outcome)

outcome は患者に対して治療を行った後の病状の変化である転帰を示す。転帰の状態は、退院時に主治医の判断に基づいた患者の状態が記載されている。表 1 に各数値の概要を示す。

#### (3) 手術情報 (opek)

opek は患者に対して行った手術情報を示す。手術情報は診療報酬点数表の診療行為の中から、手術および処置の領域に割り当てられた K コードによって符号化され、患者毎に最大 5 件の手術情報が存在する。

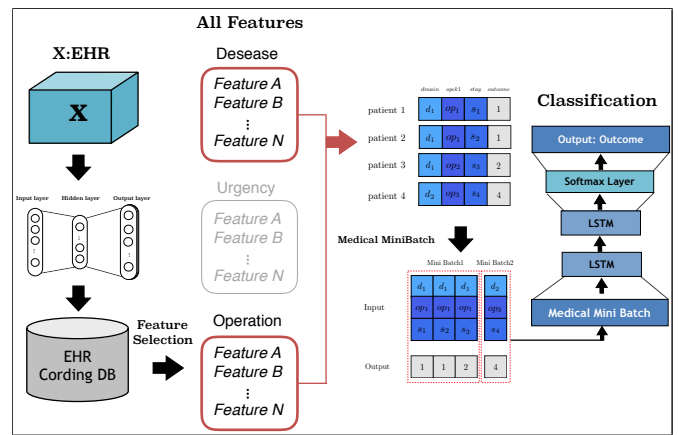


図 1 提案モデルの処理概要

### 4.2 提案モデル

本研究では、電子カルテ医療情報の分類器を構成するため長期間に渡って過去情報を反映可能な長短期記憶(Long Short Term Memory:LSTM)[4] を活用する。LSTM は RNN の発展型であり、長期記憶を可能にしたネットワークである。LSTM の構造は、RNN の中間層を情報の破棄および保持を担う入力ゲート (Input Gate), 忘却ゲート (Forget Gate), 出力ゲート (Output Gate) に置換したネットワークである。次に各ゲートの定義を (2), (3), (4) 式に示す。

$$g_j^{I,t} = f\left(\sum_i w_{ji}^{I,in} x_i^t + \sum_{j'} w_{jj'}^I z_{j'}^{t-1} + w_j^I s_j^{t-1}\right) \quad (2)$$

$$g_j^{F,t} = f\left(\sum_i w_{ji}^{F,in} x_i^t + \sum_{j'} w_{jj'}^F z_{j'}^{t-1} + w_j^F s_j^{t-1}\right) \quad (3)$$

$$g_j^{O,t} = f\left(\sum_i w_{ji}^{O,in} x_i^t + \sum_{j'} w_{jj'}^O z_{j'}^{t-1} + w_j^O s_j^t\right) \quad (4)$$

上記の (2), (3), (4) 式に示したゲートにて  $f$  はロジスティックシグモイド関数を示す。これによって各ゲートの値域は  $[0, 1]$  となる。

LSTM は 3 種のゲートの値を制御する重み値を学習することによって RNN における勾配消失問題を克服している。次に LSTM の出力値である  $z_j^t$  の順伝搬計算結果を (5) 式に示す。

$$z_j^t = g_j^{O,t} f(g_j^{F,t} s_j^{t-1} + g_j^{I,t} f(u_j^t)) \quad (5)$$

(5) 式は、(1) 式に示した RNN の出力値を各ゲートによって制御していることを示している。また (5) 式から、忘却ゲート ( $g_j^{F,t}$ ) が 0 であり入力ゲート ( $g_j^{I,t}$ ) が 1 であれば入力値を保持し続ける RNN と同値の結果が得られることが分かる。

分類器への入力となる医療情報は ICD10 コード及び K コード等によって記載されている多値論理関数であり直接的に扱うことは難しい。多値論理関数を扱う場合、一般的に複数の 2 値変数を組み合わせた表現技法である 2 進符号化 (binary coding) またはワンホット符号化 (One-hot coding) が用いられる。しかし、両手法では次元数の増大に加えて、符号化後の特徴に対して演算を行うことが難しく特徴選択を行うための類似度行列を作ることができない。そこで本研究では電子カルテ情報の符号化方法として深層学習モデルである Word2Vec Skip-Gram<sup>\*2</sup>を用いる。Word2Vec によって電子カルテの多次元情報を任意次元のベクトル空間に射影することで、各医療情報の類似度の算出が可能になる。また本研究では電子カルテの各情報を 200 次元のベクトルとして表現した。

本研究で用いたモデルである図 1 について述べる。本研究では RNN の発展系である LSTM を用いて符号化した入力情報から転帰の分類を行うためモデルの出力層にはソフトマックス関数 (Softmax Function) を分類対象クラス数分設置した。ソフトマックス関数は出力層全てのニューロンの出力値の総和が 1 の分類確率として扱えるため、最も出力値の大きいニューロンを分類結果のクラスとして判定する。また深層学習を用いたモデルでは、用途に沿った損失関数 (Error function) を設定する必要がある。そこで多クラス分類で一般的に用いられる交差エントロピー (Cross Entropy) を用いた。また図 1 の特徴選択については 4.3 節で述べる。モデルの学習に伴い、入力データを複数の Batch に分割した Mini Batch を作成する。また本研究で用いる RNN は一般的に時系列データの処理に用いられるネットワークであるが、入力情報を同一のカテゴリに分類した後、符号化を行うことで過去情報を加味した出力結果を得た研究が報告されている [5]。

本研究では符号化した入力情報をカテゴリ別に分類し Mini Batch を生成する。生成した Mini Batch を Medical Minibatch とし、生成過程を図 1 に示す。また図 1 の生成過程を事例を用いて説明する。まず各患者毎に基準とする属性によって分類するため、今回は主病名である dmain によって分類する。分類後に取得した他の属性情報である dmain, opek1, stay を入力、転帰を示す outcome を出力とした Batch データを学習データ及び分類精度を評価するためのテストデータに分割し出力する。また Medical Minibatch の生成に伴って 1 度のみ現れた入院情報を除外した。なお Medical Minibatch はモデルの学習に用いる学習セット及び評価に用いるテストセットに分割する。またモデルの学習方法として学習の進行に伴い学習係数を変化させる手法である Adagrad を適用した。また深層学習で

は多層に渡ってニューロンを設置するためモデルの表現能力は向上するが、一方で学習過程で計算した誤差関数の最小値が極小値である場合、学習誤差と汎化誤差が乖離した過学習 (Over fitting) 状態となる。そのため本研究ではモデルの過学習を抑制するためニューロンを確率的に選別し自由度を抑制する手法である Dropout およびモデルの結合重みの二乗和に正則化係数を乗じた値を重みとして用いる L2 正則化を適用した。

#### 4.3 特徴選択法

本研究では、電子カルテ医療情報の分類に有用な特徴を選択する。そこで特徴選択法として広く用いられている Wrapper に基づく手法を適用する。また電子カルテの病名や手術方法等の医療情報が単体で分類精度に影響するとは考えにくい。そのため電子カルテ医療情報の各特徴間で生じる分類結果への共起影響を考慮した特徴選択法を提案する。

##### (1) Top-K

Top-K は Wrapper 法に基づくアルゴリズムであり、電子カルテに含まれている全入力属性をそれぞれ提案モデルへの入力とし分類精度が高い属性を K 個用いて分類を行う。

##### (2) Multi Group Feature Selection(MGFS)

MGFS は Filter 法に基づくアルゴリズムであり、電子カルテ医療情報の各属性に対する Score を算出し分類器への入力を決定する。まず MGFS では (6) 式に基づき類似度行列  $\mathbf{W}_{i,j}$  を作成する。

$$\mathbf{W}_{i,j} = \mathbf{x}_i^T \mathbf{x}_j \quad (6)$$

$x_i, x_j$  は、それぞれ p 近傍距離内の符号化した電子カルテ情報であり文献 [2] に基づきドット積を利用する。また類似度行列から次数行列  $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{i,j}$  を定義する。次数行列は類似度行列の対角成分の総和であり、類似度行列との差によってグラフラプラシアン行列  $\mathbf{L}$  が定義される。以上より電子カルテ情報行列  $\mathbf{X}_{i,j}$  の多様体学習に基づく次元削減表現は (7) 式の固有値問題に帰着する。また下記の固有値問題は正規化カット最小化問題と等しい。

$$\mathbf{L}\mathbf{y} = \lambda\mathbf{D}\mathbf{y} \quad (7)$$

(7) 式にて  $\lambda$  はグラフラプラシアン行列の固有値であり、 $\mathbf{y} = [y_1, y_2, \dots, y_k]$  は上記固有値問題に対応する固有ベクトルである。本研究では採用する固有ベクトル数が出力クラス数と等しくなるように設定する。また文献 [2] では、 $\mathbf{y}$  に対して L1 正則化である Lasso 回帰を適用することで回帰係数行列  $\mathbf{a}_k$  が疎行列となり各種タスクに有用なパラメータ行列を取得できることを示している。しかし、本研究で扱う電子カルテ情報は各属性に対して複数の特徴が内包さ

\*2 <https://code.google.com/archive/p/word2vec/>

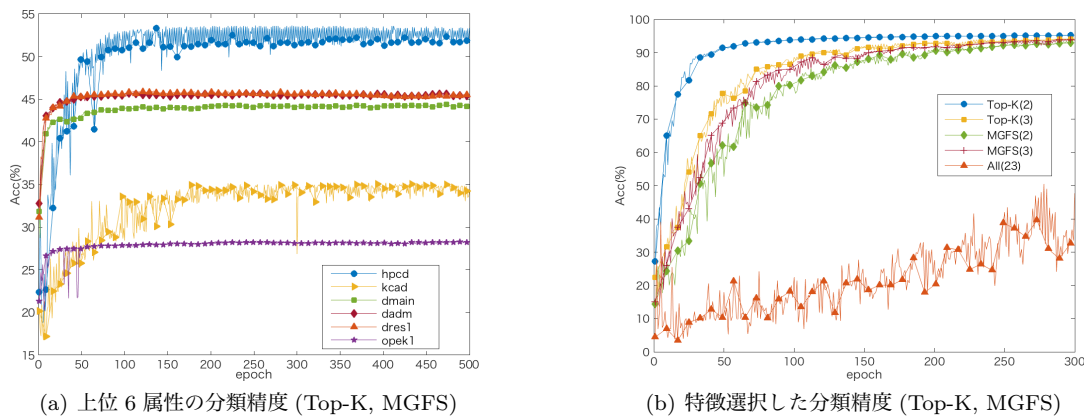


図 2 選択した特徴を利用した転帰分類精度

れており, Lasso 回帰で特徴を選定することは難しい. 例えば, 電子カルテの情報の一つに利用者が診断された主病名に関する情報がある. この病名情報は約 7000 種類に及ぶ. そのため各属性毎の特徴を評価するため, Group Lasso 推定 [6] を行う. (8) 式に Group Lasso の推定式を示す.

$$\arg \min_{\mathbf{a}_k} \left\| \mathbf{y}_k - \sum_k X^T \mathbf{a}_k \right\|^2 + \beta \sum_k \|\mathbf{a}_k\|_2 \quad (8)$$

$$\text{MGFS}(j) = \arg \max_k |\mathbf{a}_{k,j}| \quad (9)$$

L1 正則化では上記の罰則項が回帰係数に対する絶対値の総和であるが, Group Lasso では L2 正則化を施すことで各説明変数のグループ毎に回帰係数行列を疎行列にする働きがある. 本研究では Word2Vec を用いて医療情報を符号化したため, 符号化後の各属性毎にグループ化を施した. また  $\beta$  はパラメータであり, 本実験では (8) 式の最小二乗誤差 (Mean Squared Error:MSE) が最小になるよう Grid Search によってパラメータを推定した. 一般的に Lasso の推定解は解析的に算出することが難しいため, ブロック座標降下法 (Block Coordinate Descent:BCD)[7] を用いて回帰係数行列  $\mathbf{a}_k$  を推定し, (9) 式に基づき, 各グループのスコアを算出する.

## 5. 評価実験

本研究では提案モデル及び特徴選択の有効性を検証するため, 実データを用いた実験を行った. 使用したデータセットは 33208 件の入院情報を持つ電子カルテ情報であり, テストデータに対する分類精度を正確に検証するため, 4 分割交差検証 (4-fold cross validation) を用いた. 実験は 16GB のメモリ, Intel Core i5 3.3GHz の CPU を搭載した iMac のマシン上で実施した. 本実験は, 以下の諸問題に取り組む.

- Q1 電子カルテ医療情報から転帰分類に有用な特徴の検討
- Q2 提案手法における分類精度と既存手法との検証

表 2 特徴選択法を用いた分類精度

	Acc(%)	次元数	属性
Top-K(2)	<b>95.1</b>	400	hpcd,dres1
MGFS(2)	94.0	400	opek1,hpcd
Top-K(3)	92.8	600	hpcd,dres1,dadm
MGFS(3)	<b>93.6</b>	600	opek1, hpcd, dmain
All Feature	47.6	4600	All feature

### 5.1 Q1: 転帰分類に有用な特徴の検討

本節では, 転帰分類に有用な特徴発見のため Top-K, MGFS を実データに適用し選択した複数属性を用いて分類精度を検証したケーススタディを示す. 図 2(a) は提案モデルに転帰属性を除いた電子カルテの全 23 属性をそれぞれ入力とし, MGFS にて選定した上位 6 属性の Top-K による分類精度を示している. なお横軸は学習回数, 縦軸は分類精度を示す. 各属性について説明する. hpcd は各医療機関を識別する情報, kcad は患者が受診した診療科情報, dadm は入院契機となった傷病, dres1 は医療資源を最も消費した傷病である. 図 2(a) において属性 opek1 は MGFS にて算出したスコアでは最上位を示したが, Top-K による分類精度の高さは全属性において  $\frac{9}{23}$  番目であった. そのため属性 opek1 単体では高い分類精度は望めないが, 他の属性を加えることで分類精度に影響を与える共起因子であると考えられる.

図 2(b) は Top-K および MGFS を用いて選定した特徴を提案分類モデルへ入力した際の分類精度を示している. 各特徴選択法の結果の詳細は表 2 に示す. より具体的には, Top-K では分類への有用性は低いと考えられる属性 opek1 を分類器への入力とした場合であっても Top-K により選択した特徴と同程度の分類精度を得た. そのため複数特徴間の共起影響が分類精度に関与することが確認できる. また Top-K および MGFS にて抽出した特徴のみを分類に使用した場合, 電子カルテに含まれる全特徴を入力とした結果と比較すると分類精度の大幅な向上が確認できる. これ

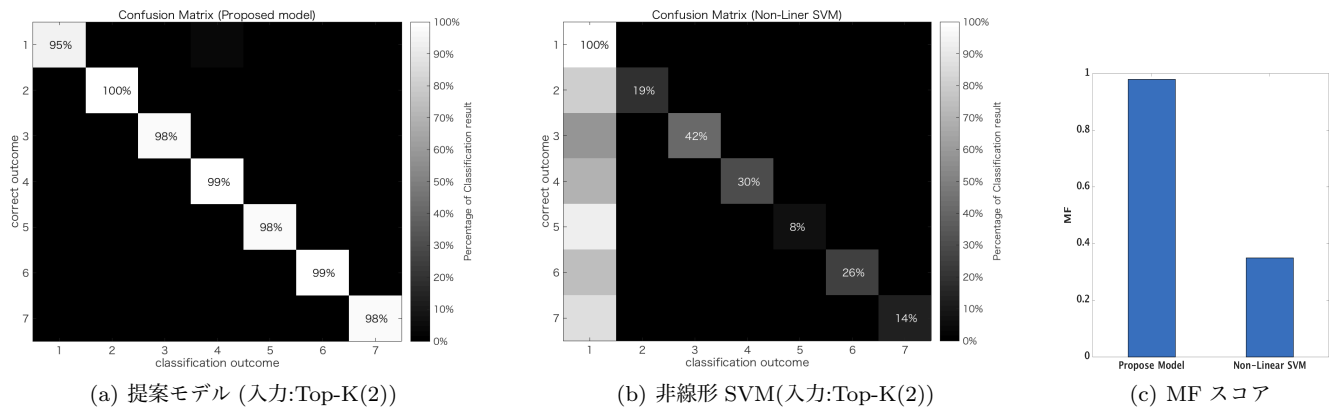


図 3 提案モデルと既存モデルの比較

は全特徴を入力とした場合の次元数が非常に多く、転帰情報を分類するための情報が不明瞭になった点に加えて、学習時間の大幅な増加が考えられ特徴を選定することの重要性を示した。

### 5.2 Q2: 提案手法の精度

提案モデルの分類精度を検証するため、既存手法である非線形 SVM との比較を行った。非線形 SVM のカーネル関数として一般的に用いられているラジアル基底関数 (Radial Basis Function:RBF) を利用した、図 3 は表 2 にて最も分類精度の高い特徴選択結果である Top-K(2) を用いた場合の提案モデルと非線形 SVM における分類の精度を示している。より具体的には分類の正解クラスおよび推定したクラスに関する混同行列 (Confusion Matrix:CM) を作成し、マイクロ平均 F 値 (Micro average F-value:MF) を (10) 式で計算した。(10) 式において  $P_i$  は各クラスの適合率、 $R_i$  は各クラスの再現率であり  $N$  はクラス数を示す。

$$MF = \frac{1}{N} \sum_i \frac{2P_i R_i}{P_i + R_i} \quad (10)$$

正確にクラスを推定することができた場合、混同行列は対角行列となり、MF=1 となる。図 3(a)、図 3(b) はそれぞれ提案モデルと非線形 SVM の混同行列を示しており、内部数値は各クラスの正答割合を示す。図 3(a) から提案モデルの分類器は、全ての転帰クラスにおいて高精度の分類が実現できていることが分かる。特に図 3(b) と比較して転帰クラス 5~7 の大幅な精度改善が確認できる。対応する転帰はいずれも利用者の傷病が悪化した場合であり他利用者の過去情報を加味したことが高精度に起因したと考えられる。また図 3(b) は転帰クラス 1 に大部分のテストデータが分類されており、他クラスの分類精度が明確に低い。図 3(c) は両手法の MF を示しており、既存モデルと比較すると提案モデル分類性能がより高精度であることが分かる。

## 6. むすび

本研究では電子カルテ医療情報の多角的解析のための医療情報分類手法として医療情報から分類に有用な特徴を特定し、深層学習を用いた分類モデルを構築した。提案手法は、既存の特徴選択法では特定の難しい分類に有用な特徴の選定が可能であり、実データ用いて転帰分類に関する実験を行った結果、既存モデルと比較して提案モデルの分類精度の向上を示した。

### 参考文献

- [1] Y.Bengio, P.Lamblim, D.Popovici, and H.Larochelle. "Greedy layer-wise training of deep networks", Conference on Neural Information Processing Systems(NIPS),pp.153-160(2006)
- [2] Deng Cai, Chiyuan Zhang, Xiaofei He."Unsupervised Feature Selection for Multi-Cluster Data", ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD),pp.333-342(2010)
- [3] Hwanjo Yu, Jadeo Vaidya, Xiaoqian Jiang, "Privacy-Preserving SVM Classification on Vertically Partitioned Data", The Pacific-Asia Conference on Knowledge Discovery and Data Mining(PAKDD), pp647-656(2006)
- [4] A.Graves, "Supervised sequence labelling with recurrent neural networks", volume of 385 of Studies in Computational Intelligence.Springer-Verlag Berlin Heidelberg(2012)
- [5] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, Domonkos Tikk, "Session-Based Recommendations with Recurrent Neural Networks", arXiv:1511.06939v4,ICLR2016
- [6] M.Yuan and Y.Lin, "Model selection and estimation in regression with grouped variables", Journal of the Royal Statistical Society:Statistical Methodology,pp49-67(2006)
- [7] Zhiwei Qin, Katya Scheinberg, Donald Goldfarb, "Efficient block-coordinate descent algorithms for the Group Lasso", Mathematical Programming Computation,Volume5,Issue2,pp143-169(2013)
- [8] Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, "Learning to read chest X-rays:recurrent neural cascade model for automated image annotation", Conference on Computer Vision and Pattern Recognition(CVPR), pp2497-2506(2016)

## 正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
3 ページ 5 行目	本研究では、京都府の複数病院より提供を受けた電子カルテ情報を用いる。	本研究では、京都大学医療経済学分野により実施されている QIP プロジェクトより提供される匿名化された医療情報を用いる。
3 ページ 6 行目	電子カルテには、病名や術式を文字や数値で表現するために ICD10 コードや K コードが用いられている。そこで ICD10 並びに K コード等によって記録されている属性を代表して dmain,outcome,opek1 のそれぞれについて述べる。また dmain 等を電子カルテ情報の属性と呼び本研究で扱う電子カルテには 24 種の属性が存在する。	QIP では全国の参加病院より匿名化された DPC データの提供を受けており、DPC データは DPC/PDPS と呼ばれる診療報酬の包括支払い制度に参加する病院で作成が必須となっている全国共通フォーマットデータである。このデータには行った診療行為データのほか、病名や術式、あるいは入退院時の状況が含まれる。これらのデータは文字列表記のほか、傷病は国際標準の ICD10 コード、手術は診療報酬で規定された K コードと呼ばれるものによりコーディングされている。そこで ICD10 並びに K コード等によって記録されている属性を代表して主病名、手術名、転帰のそれぞれについて述べる。またここではこれらの項目を電子カルテ情報の属性と呼び、本研究では 24 種の属性を利用した。