

カーネルロジスティック回帰を用いた C型慢性肝炎の肝線維化ステージ推定

大崎 美穂^{1,a)} 松田 健司¹ ワン ペン¹ 片桐 滋¹ 横井 英人² 高林 克日己³

受付日 2014年12月31日, 採録日 2015年08月12日

概要: 従来, C型慢性肝炎の病状把握と治療のため, 血液・尿の臨床検査の結果から肝線維化ステージを推定する試みがなされてきた. 過去の研究の多くは分類器に重回帰 (MLR) やロジスティック回帰 (LOGR) を用いており, 推定性能はあまり高くなかった. サポートベクターマシン (SVM) により推定性能を向上した研究もあるが, SVM にはクラスの事後確率が得られない, 多クラスへの拡張が容易ではない問題があった. そこで本研究では, LOGR と SVM の利点をあわせ持つカーネルロジスティック回帰 (KLOGR) の適用を提案し, 従来適用された分類器と推定性能を比較する実験を行い, KLOGR の有効性を検証した. 実験では, 肝線維化の深刻さの2クラス問題, 肝硬変の有無の2クラス問題の各々について, 分類器ごとに Leave-One-Out 交差検定で推定性能を導出した. 結果として, KLOGR は SVM 等の他の分類器よりも高い推定性能を示した. 特に, 従来最もよく使われてきた MLR や LOGR よりも, 推定性能が 8.32~12.50% 高かった. クラスの事後確率の導出, 多クラスへの拡張の容易さという利点も含め, 肝線維化ステージ推定における KLOGR の有効性が明らかになった.

キーワード: C型慢性肝炎, 肝線維化ステージ推定, カーネルロジスティック回帰, サポートベクターマシン

Prediction of Liver Fibrosis Stages in Chronic Hepatitis C Using Kernel Logistic Regression

MIHO OHSAKI^{1,a)} KENJI MATSUDA¹ PENG WANG¹ SHIGERU KATAGIRI¹ HIDETO YOKOI²
KATSUHIKO TAKABAYASHI³

Received: December 31, 2014, Accepted: August 12, 2015

Abstract: For diagnosis and treatment of chronic hepatitis C, there have been studies to predict the stage of liver fibrosis using the results of clinical blood and urine tests. Many conventional studies used multiple linear regression (MLR) and logistic regression (LOGR) as classifiers and could not achieve high prediction performance. There exist a few studies using support vector machine (SVM) which achieved the better prediction performance. However, SVM has drawbacks that it does not derive the posterior probabilities of classes and is not easily applicable to multiclass problems. We propose to apply kernel logistic regression (KLOGR) which possesses the advantages of LOGR and SVM and evaluate KLOGR through experiments to compare it with the conventionally used classifiers. For each classifier, we estimated the prediction performance by the leave-one-out cross-validation in two-class problems, one on the seriousness of fibrosis and the other on the existence of cirrhosis. As a result, KLOGR constantly outperformed other competitive classifiers including SVM. Especially, KLOGR achieved 8.32 to 12.50% higher prediction performance than that of MLR and LOGR, which have been most frequently used. With this result and the intrinsic advantages of KLOGR, the effectiveness of KLOGR in liver fibrosis stage prediction was clarified.

Keywords: chronic hepatitis C, prediction of liver fibrosis stages, kernel logistic regression, support vector machine

1. はじめに

C型肝炎ウイルスに感染すると肝臓が炎症を起し、様々な肝機能障害と肝臓の硬化が生じる。ウイルス感染と肝機能障害が6カ月以上持続する病態は、C型慢性肝炎と呼ばれる。C型慢性肝炎には10~30年を経て肝硬変や肝がんに至る恐れがあり、我が国では150万人を超える感染者が存在するといわれている[22]。この病状把握と治療のためには、慢性的な肝細胞の破壊と再生を経て肝臓が硬化する肝線維化の度合い（肝線維化ステージ）を把握する必要がある。

肝線維化の診断方法としては、直接採取した肝細胞を医師が顕微鏡観察し、F0~F4の5段階ステージ分類を行う肝生検が主流である。しかし、この方法は侵襲的でコストやリスクの問題があるため、代替手段として肝線維化ステージ推定が試みられてきた[18], [26], [27], [31]。なお、ここでいうステージ推定とは、顕微鏡観察以外の間接的な情報を用いる推定システムが「もし肝生検を行ったならば、医師が分類するであろうステージ」を推し量ること、あるいは、医師が推し量る補助をすることである。

肝線維化ステージ推定には大別して次の2つの方法がある。1つは、ステージ推定を行う分類器に血液・尿の検査値を入力し、医師の代わりに分類器がステージの分類結果を出力する方法である[15], [17], [20], [36]。もう1つは、肝臓に照射した超音波の反響を映像化し、医師がその映像を見てステージ分類する方法である[7], [21], [23], [24], [33]。現状では、どちらの方法の推定性能が高いかに関しては諸説あり、方法の間には相補的な利点と欠点が存在する[7], [21], [23], [24], [33]。推定性能の向上には、各方法の改善と方法の組合せによる改善のいずれもが重要であろう。本研究では特に、血液・尿の検査値と分類器を用いた肝線維化ステージ推定の改善に主眼を置く。

従来研究の多くは、分類器として重回帰（MLR: Multiple Linear Regression）やロジスティック回帰（LOGR: Logistic Regression）を用いていたが、推定性能が十分に高いとはいえない。ただし、LOGRはクラスの事後確率が得られる利点、多クラスへの拡張が容易である利点を持つ[18], [26], [27], [31]。サポートベクターマシン（SVM: Support Vector Machine）を適用した研究では、比較的高い推定性能が得られたが、SVMにはクラスの事後確率が得られない、多クラスへの拡張が容易ではないという問

題がある[20]。

従来研究には分類器の問題に加えて評価方法の問題もある。具体的には、推定に対する検査値の特微量の効果と分類器の効果を区別していないこと、各研究がそれぞれ異なる独自のデータと条件設定を用いたために研究の間で推定性能を比較できないことがあげられる。

以上の背景より、本研究ではLOGRとSVMの利点をあわせ持つカーネルロジスティック回帰（KLOGR: Kernel Logistic Regression）の適用を提案し、その有効性を実験検証する[25]。実験に際しては、結果の一般性を確保するようにデータの選定と条件設定を行うとともに、分類器の間で共通の特微量を用いることで特微量の効果を公平にして分類器を評価する。対象のデータセットから一般的な検査項目を選び出し、これらの検査項目からなり、かつ、分類器の間で共通する特微量を生成する。この特微量を用いた実験を行い、従来適用された分類器と比較することでKLOGRの有効性を明らかにする。

本論文の構成は以下のとおりである。2章では、従来の肝線維化ステージ推定の試みを説明する。3章では、従来研究で用いられた分類器と対比しながら、本研究で適用を提案するKLOGRについて説明する。4章では、様々な分類器の推定性能を比較し、KLOGRの有効性を検証した実験について述べる。5章では、本論文のまとめと今後の展望を示す。

2. 従来の肝線維化ステージ推定

肝線維化ステージ推定に関する要因として、対象問題や、特微量、分類器、推定性能の評価基準をあげることができる。現在、肝線維化の深刻さ（F0-F1/F2-F4）を分類する2クラス問題と肝硬変の有無（F0-F3/F4）を分類する2クラス問題が一般的な対象問題である。

推定システムは前処理、特微量抽出、分類の3段階からなり、これらの設計結果は推定に大きく寄与する。従来研究の間では前処理は共通していることが多いが、特微量、分類器、データ、条件設定が個々に異なっていた[15], [17], [18], [20], [26], [27], [31], [36]。このため、個々の研究で得られた推定性能に何が寄与したのかを特定できないうえに、研究の間で推定性能を比較できない問題があった。推定性能（分類器が出力するクラスは、肝生検を行った場合に得られる正しいクラスとどの程度一致するか）の評価では、医療データ特有の性質を考慮して複数の評価基準を用いることが多い。

以下では、従来研究における前処理と特微量抽出、分類器、推定性能の評価基準の各々について、さらに詳しく述べる。従来の前処理では、検査値に特に変換を加えない、あるいはヒューリスティクスに基づき一部の検査値を対数変換していた。特微量抽出では、肝線維化ステージを知りたい時点の前処理済みの検査値を検査項目ごとに並べ、得られたベク

¹ 同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University, Kyotanabe, Kyoto 610-0321, Japan

² 香川大学医学部附属病院
Kagawa University Hospital, Kita-gun, Kagawa 761-0793, Japan

³ 千葉大学医学部附属病院
Chiba University Hospital, Chiba 260-8677, Japan

a) mohsaki@mail.doshisha.ac.jp

トルを特徴量としていた。特徴量に用いる検査項目としては、病状のモニタリング用に測定される通常の血液・尿検査や、肝硬変や肝がんの可能性が疑われる場合に実施する精密検査があげられる [15], [17], [18], [20], [26], [27], [31], [36].

精密検査を行えば病状に関する詳細な情報が得られるため、当然、推定性能が高くなることが多い [15], [17], [20], [36]. しかし、特徴量や分類器の改善により、普及しておりコストが低い通常の血液・尿検査で高い推定性能を得ることも有益であろう。また、特徴量や分類器の改善は血液・尿検査に限らず、精密検査を用いた推定にも貢献しうる。

意思決定をとともなう分野、特に生命に関わる分野では一般に、推定結果に対する確信度の目安としてクラスの事後確率を重視している [4], [30]. また、現在は推定の難しさゆえに2クラス問題が主流であるが、将来的にF0-F4を推定する5クラス問題を扱うことも考えられる。これらの点をふまえ、分類器にはMLRやLOGRを用いた従来研究が多かった [15], [17], [36].

LOGRは原理的に、推定結果とともにクラスの事後確率（肝線維化ステージ推定の場合はステージの事後確率）を導出でき、多クラスへの拡張が容易という利点も持つ。MLRは、LOGRのようにクラスの事後確率を直接導出するものではなく、多クラスへの拡張も容易とはいいがたいが、仕組みの簡潔さゆえに用いられることが多かった。しかし、これらは線形の識別境界しか表現できないため、自ずと推定性能は限られる。

近年ではSVMの適用も見られ、その非線形な識別境界の表現能力と幾何マージン最大化に基づく汎化能力により、より高い推定性能が得られている [20]. その反面、SVMにはクラスの事後確率が得られないという問題がある。SVMはクラスの事後確率の推定ではなく、クラスの推定性能の向上を目指して学習する。ゆえに、形式的には学習結果から事後確率に類する量を導出できるが、厳密な意味でのクラスの事後確率は得られない。また、多クラスへの拡張も容易ではない。SVMの多クラス拡張版は存在するが [11], その評価はまだ定まっておらず、2クラス用のSVMを多クラスに適用する場合はペアワイズ等の手続きが必要となる。

推定性能の評価用の基準には、AUC (Area Under the Curve of a receiver operating characteristic) や、Acc (Accuracy), Sens (Sensitivity), Spec (Specificity), PPV (Positive Prediction Value), NPV (Negative Prediction Value) がある [1], [2], [3], [26], [27]. ただし、AUCは他の評価基準とは異なり、パラメータ設定に対する分類器の頑強性を表す。データに基づく客観的なパラメータ最適化が難しい場合にAUCは必要であるが、最適なパラメータ設定下の推定性能の評価には向かない [14]. 本研究は最適なパラメータ設定下の推定性能に主眼を置くので、これ以降はAUCには触れない。

Acc, Sens, Spec, PPV, NPVの定義を式(1)に示す。式中、症状が重いクラスは陽性、 N は全データ数、 N_{TP} , N_{TN} , N_{FP} , N_{FN} は、各々、真陽性 (True Positive), 真陰性 (True Negative), 偽陽性 (False Positive), 偽陰性 (False Negative) のデータ数である。多面的な評価を意図して5つの評価基準とも重視する、あるいは、陽性・陰性の個別の推定性能を見ることを意図してAcc以外の4つの評価基準を重視することが多い。優先順位を付ける場合は、真陽性と真陰性の取りこぼしを抑えつつ (SensとSpecを高くしつつ)、陽性を正確に判定する (PPVを高くする) という考えが一般的である。

$$\begin{aligned} \text{Acc} &= \frac{N_{TP} + N_{TN}}{N} & (1) \\ \text{Sens} &= \frac{N_{TP}}{N_{TP} + N_{FN}} & \text{Spec} = \frac{N_{TN}}{N_{TN} + N_{FP}} \\ \text{PPV} &= \frac{N_{TP}}{N_{TP} + N_{FP}} & \text{NPV} = \frac{N_{TN}}{N_{TN} + N_{FN}} \end{aligned}$$

Acc, Sens, Spec, PPV, NPVの間にはトレードオフの関係がある。たとえば全事例を軽症に分類すると、Specの値は100%となるが、Sensの値は0%、PPVの値は定義不能となる（重症患者を見落とすリスクが大きくなる）。また、C型慢性肝炎の検査履歴では軽症に比べて重症事例が少ないため、この傾向によりAcc, NPVの値は高くなる。全事例を重症に分類しても同様の問題が起こる。従来研究では、このトレードオフとC型慢性肝炎の検査履歴に存在するクラス間のデータ分布の偏りを考慮し、複数の評価基準による多面的な評価を行っていた。

多面的な評価はとても重要であるが、トレードオフ関係を持つ評価基準の値を個別に検討するのみでは不十分であると思われる。個別の検討では、ケースバイケースの議論に発散する問題、評価基準の間のバランスをとらえる視点が欠ける問題が生じるからである。また、異なる特徴量や分類器を比較選択したいとき、多様な複数の評価基準はこれを困難にしてしまう。ゆえに、複数の評価基準のトレードオフを包括した総合的な評価基準が望まれる。

推定システムには未知データに対する汎化能力が求められる。この汎化能力を見積もるため、訓練に用いたデータとは異なるデータに対して評価基準の値を得て推定性能を評価する。従来、データセットを訓練用とテスト用に2分割する簡便な手続きが採用されていたが、より厳密な見積りが可能な手続きを検討すべきであろう。

ここで、残された課題を整理する。主問題：肝線維化ステージ推定の性能は十分高いとはいいがたく、特徴量と分類器の改善による性能向上が必要である。副問題1：研究間でデータ・条件設定が異なるため、推定性能の比較が困難である。副問題2：推定に対する特徴量と分類器の効果が明確に区別されていない。副問題3：推定性能の評価基準が複数あるため、総合的な評価が難しい。

3. カーネルロジスティック回帰の適用の提案と実験設計

3.1 従来適用された分類器とカーネルロジスティック回帰

主問題の解決, すなわち, 肝線維化ステージ推定の性能向上に向け, 我々は KLOGR を適用することを提案する [25]. 3.1 節では, 従来研究で用いられた分類器のうち KLOGR に深い関連がある SVM と対比しながら, KLOGR を説明する. また, 分類器を実装するうえで重要な事項 (バイアス項の表現, 正則化におけるバイアス項の有無, カットオフの値の設定) も検討する. バイアス項とカットオフは同義で使われることもあるが, 本研究ではこれらを次のように定義して区別する. バイアス項は学習中に設定される識別関数の切片, カットオフは学習後に設定される識別の閾値 (バイアス項を補正したものともいえる).

従来研究では, MLR, LOGR, SVM を適用している. また, 比較対象として学習過程を含まない最近傍法 (1-NN: One-Nearest-Neighbor) を適用し, その推定性能を学習過程を含む分類器が超えるべき下限値と見なしている. 1-NN, MLR, LOGR の説明は省略するが, これらに関しては解説 [6] や適用例 [15], [17], [20], [36] を参照されたい. なお, MLR では, 陰性ならば 0, 陽性ならば 1 の値をとる 1 つの出力変数を予測対象とした. 以下では SVM と KLOGR を詳しく説明し, 両者の関係を示す. 式中の記号は表 1 にリスト化して示すとおりである. なお, すべての分類器は式 (2) に示す共通の分類規則を持つ.

$$C(\mathbf{x}) = C_k, \quad \text{iff } k = \underset{j}{\operatorname{argmax}} g_j(\mathbf{x}) \quad (2)$$

サポートベクターマシン (SVM) [10], [12] は, クラス境界とその近傍に位置するデータ点の距離である幾何マージンを最大化することで, 未知データへの汎化を目指した線形分類器である. オリジナルの SVM は, クラスの所属を指示関数 t_n で表現した 2 クラス問題を対象とする. 非線形な識別境界の推定を可能にするため, SVM ではカーネル関数の導入が一般的である [10], [12].

SVM では, 式 (3) のようにカーネル関数の加重和で 2 クラスの境界を決定する関数 y を表す. y の符号と指示変数の符号が一致する, すなわち, 正分類する (ただし, 若干の誤分類は許容する) という制約を立てる.

$$y = \sum_{m=1}^N \alpha_m \mathcal{K}(\mathbf{x}, \mathbf{x}_m) + b \quad (3)$$

この制約と, クラス境界と最近傍の事例との距離である幾何マージンの最大化という目的を集約すると, 式 (4) に示す目的関数が得られる. ただし, ξ_n は誤分類の許容度合いを表すスラック変数, c は誤分類を重視する度合いを表すハイパーパラメータである. $J(\boldsymbol{\alpha}, \boldsymbol{\xi}, b)$ を最小化すると, 式 (4) の第 1 項は誤分類の度合いの総和を小さくすること,

表 1 3 章の式における記号とその説明

Table 1 List of symbols and their meanings used in the equations in Section 3.

K :	クラスの数.
N :	訓練データの事例の数.
\mathbf{x} :	特徴量ベクトル.
\mathbf{x}_n :	訓練データの n 番目の事例が持つ特徴量ベクトル.
$\mathcal{K}(\mathbf{x}, \mathbf{x}_n)$:	\mathbf{x} と \mathbf{x}_n の類似度を表すカーネル関数. 本研究では, 最も一般的なガウスクーネル $\mathcal{K}(\mathbf{x}, \mathbf{x}_n) = \exp(-\ \mathbf{x} - \mathbf{x}_n\ ^2/2\sigma^2)$ とする [6].
Φ :	グラム行列. 訓練データの事例の全組合せについて求めたカーネル関数値を要素に持つ.
C :	クラス.
C_k :	k 番目のクラス.
$\boldsymbol{\alpha}$:	双対問題におけるパラメータベクトル. このパラメータ数は訓練データの事例数に等しい.
$\boldsymbol{\alpha}_k$:	双対問題における k 番目のクラスに対応するパラメータベクトル. 要素として α_{km} ($m = 1, 2, \dots, N$) を持つ.
b :	双対問題におけるバイアス項.
b_k :	双対問題における k 番目のクラスに対応するバイアス項.
$g(\mathbf{x})$:	識別関数.
$g_k(\mathbf{x})$:	k 番目のクラスに対応する識別関数.
t_n :	2 クラス問題において, 訓練データの n 番目の事例がどちらのクラスに属するかを示す指示関数. 一方のクラスに属するときに +1, 他方のクラスに属するときに -1 と取る.
t_{nk} :	k クラス問題において, 訓練データの n 番目の事例が k 番目のクラスに属することを示す指示変数. k 番目のクラスに属するときに 1, 属さないときに 0 と取る.

第 2 項は幾何マージンを大きくすることに相当する. 第 2 項から, 幾何マージンの最大化は形式的にパラメータの L^2 ノルム正則化と等しいことが分かる.

$$J(\boldsymbol{\alpha}, \boldsymbol{\xi}, b) = c \sum_{n=1}^N \xi_n + \frac{1}{2} \boldsymbol{\alpha}^T \Phi \boldsymbol{\alpha} \quad (4)$$

subject to: $t_n y_n \geq 1 - \xi_n, \xi_n \geq 0$

訓練段階では, 式 (4) の最小化によりパラメータ $\boldsymbol{\alpha}$ と $\boldsymbol{\xi}$ が求まる. パラメータの学習には, 事例対の選択と事例対に対応するパラメータの値の導出を繰り返す SMO アルゴリズムが用いられることが多い. バイアス項 b は双対問題上で直接は解けないため, 主問題に立ち返って導出する方法が提案されている [6], [12]. 本研究では, これらの標準的な導出方法のうち, 式 (5) を用いる方法を採用する (文献 [12] の Remark 6.4 参照). ただし, $\phi(\mathbf{x}_n)$ は無限次元の双対空間への写像関数であり, その内積はカーネル関数 $\mathcal{K}(\mathbf{x}, \mathbf{x}_n)$ に等しい. また, SVM では一般的とはいいがたいが, MLR, LOGR, KLOGR と同様にバイアス項の代わりにカットオフを導入する方法も検討する. カットオフの値を求める手続きは全分類器に共通するので, 3.2 節の末尾にまとめて説明する.

$$b = -\frac{1}{2} \left(\max_{t_n=-1} (\mathbf{w}^T \phi(\mathbf{x}_n)) + \min_{t_n=1} (\mathbf{w}^T \phi(\mathbf{x}_n)) \right) \quad (5)$$

$$\text{where } \mathbf{w} = \sum_{n=1}^N \alpha_n t_n \phi(\mathbf{x}_n)$$

テスト段階では、式 (6) を式 (2) に代入することで、いい換えると、式 (3) の符号を判定することで、テスト事例にクラスを割り当てる。

$$g_{-1}(\mathbf{x}) = 0, \quad g_{+1}(\mathbf{x}) = y \quad (6)$$

LOGR [9] は、線形回帰式を用いてクラスの事後確率を推定し、それに基づきクラスを推定する分類器である。元々は2クラス問題を対象としていたが、簡単な式の置き換えのみで多クラスにも拡張可能である。KLOGR [32], [38] はLOGR にカーネル関数を導入し、SVM と同様に非線形な識別境界を可能にした手法である。

KLOGR では、クラスの所属を指示関数 t_{nk} で表現する k クラス問題を扱う。式 (7) のようにカーネル関数の加重和で各クラスに対応する関数 y を定義する。そして、式 (7) を式 (8) のソフトマックス関数に代入し、その値を各クラスの事後確率と見なす。

$$y_k = \sum_{m=1}^N \alpha_{mk} \mathcal{K}(\mathbf{x}, \mathbf{x}_m) + b_k \quad (7)$$

$$p(C_k | \mathbf{x}) = \frac{\exp(y_k)}{\sum_{l=1}^K \exp(y_l)} \quad (8)$$

学習では式 (9) の目的関数の最小化を図る。式 (9) の第1項は、推定したクラスの事後確率とデータとの整合度合いを表す交差エントロピー誤差関数である。第2項は、過学習を防止し、かつ、リプレゼンタ定理を満たすように加えたパラメータの L^2 ノルム正則化項である。これは形式的に幾何マージン最大化と同じ効果を生む。式 (9) の λ は、正則化項に対する重みを表すハイパーパラメータである。この λ は式 (4) の c と逆の関係にある。訓練段階では、式 (9) の最小化によりパラメータ α が求まる。目的関数が凸であるため、パラメータの学習には勾配に基づく最適化手法を用いることが多い。

KLOGR のバイアス項 b_k の設定には、目的関数にバイアス項を含む方法 [8], [19], [34] と含まない方法 [19], [28], [32], [38] がある。バイアス項を含む方法では、ダミーのカーネルを追加してグラム行列 Φ を拡張し、 b_k をダミーのカーネルの係数 α_{0k} と見なして α に埋め込む。そして、他のパラメータと同じ手続きでバイアス項の値を求める。バイアス項を含まない方法では、式 (7) の b_k を 0 に設定して目的関数を立て、パラメータ α とは別途 b_k を求める。 b_k の求め方としては、生物医学分野でよく使われる手続き（学習後にバイアス項をカットオフと見な

して検証データで設定する) [26], [27] があげられる。

本研究では、目的関数が SVM と同様になることと生物医学分野で一般的な手続きを考慮し、バイアス項を含めずにカットオフを導入する方法を採用する。なお、カットオフの導入は本研究で用いる分類器に共通するので、3.2 節の最後にまとめて詳細に議論する。

$$J(\alpha_1, \alpha_2, \dots, \alpha_K) \quad (9)$$

$$= -\sum_{n=1}^N \sum_{k=1}^K t_{nk} \ln p(C_k | \mathbf{x}_n) + \frac{\lambda}{2} \sum_{k=1}^K \alpha_k^T \Phi \alpha_k$$

テスト段階では、式 (10) を式 (2) に代入することで、いい換えると、式 (10) の値が最大となるクラスを求めることで、テスト事例にクラスを割り当てる。

$$g_k(\mathbf{x}) = p(C_k | \mathbf{x}) \quad (10)$$

式 (4) と式 (9) を比較すると、KLOGR は SVM と同様に、カーネル化による非線形構造の記述能力と幾何マージン最大化による汎化能力を持つと分かる。また、式 (8) の形から、KLOGR は LOGR と同じく、クラスの事後確率を導出可能であり、多クラスへの拡張の容易さも持っている。このように KLOGR は SVM と LOGR の利点を兼ね備えており、我々は KLOGR が肝線維化ステージ推定に役立つと考えた。

3.2 分類器の実装方法、ハイパーパラメータとカットオフの設定方法

分類器の理論的な枠組みに加え、分類器の実装方法も推定性能に影響を及ぼしうる。特に、バイアス項を表現する、正則化の計算でバイアス項を扱う、ハイパーパラメータの値を決定する、カットオフの値を決定する、という4つの事項は重要である。

バイアス項の表現には、識別関数にバイアス項を明示的に加える方法、つねに1の値をとる入力変数を追加して特徴量ベクトルを拡張し、非明示的に埋め込む方法がある。後者はバイアス項をパラメータベクトルに含むため、前者よりも定式化における利便性が高い。また、特徴量ベクトルを構成する入力変数の単位やレンジがパラメータとバイアス項の値に影響するので、この影響を取り除く必要がある。これらをふまえて本研究では、正規化した拡張特徴量ベクトルを全分類器に共通して用いる。

特徴量空間から無限次元の双対空間への写像をとるカーネル分類器 (SVM, KLOGR) では、特徴量空間のバイアス項とは別に、双対空間のバイアス項も考える必要がある。SVM では、双対空間のバイアス項を求める方法が確立しているため、これに従う [12]。また、KLOGR 等との比較のため、カットオフを用いる方法も試行する。KLOGR ではカットオフがバイアス項の役割を果たすため、双対空間上のバイアス項は導入しない。

正則化をとまなう分類器の場合、正則化項のノルム計算にバイアス項を含める方法と含めない方法がある。前者では識別境界のシフト範囲が制約され過ぎる恐れがあり、後者の方が一般的に用いられる [6], [16]。ゆえに本研究では、正則化をとまなう分類器に共通してノルム計算にはバイアス項を含めないことにする。

ハイパーパラメータの値は分類器の性能に大きな影響を及ぼす上に、その探索の決定基準は非凸であり、値の設定が難しい。この理由より、値域内を漏れなく等間隔に調べる単純なグリッドサーチが採用されることが多い。決定基準には目的に応じて Acc, Sens, Spec, PPV, NPV の一部や組合せが用いられ、決定基準の値は過学習を防ぐために訓練データとは異なる検証データにより算出される。本研究もこれに習い、ハイパーパラメータを持つ分類器には上記の決定方法を採用する。

ここで、カットオフの意義と設定方法について述べる。前述したように、生物医学分野では慣習的に、クラスの事後確率を推定するタイプの分類器に対し、学習後にカットオフを導入することが多い [26], [27]。カットオフの導入は、学習を通して推定したクラスの事後確率を分類器の性能に基づき補正することを意味するが、これを明確に意識した研究は少ない [16]。このように、生成モデル/識別モデルアプローチ（クラスの事後確率推定に主眼を置く考え方）から識別関数アプローチ（クラスの正分類に主眼を置く考え方）[6]に変更する場合は、注意が必要である。カットオフで補正したクラスの事後確率は厳密には確率の定義から外れているので、このことを認識しておくべきである。

カットオフは以下の理由から有効である可能性が高い。生物医学分野では、カットオフの有効性を示す事例が多く見られる [26], [27]。機械学習分野では、生成モデル/識別モデルアプローチの事前学習と識別関数アプローチの再学習の組合せが有効であるといわれている [13]。2クラス問題においてカットオフは識別関数のバイアス項に相当するが、バイアス項を他のパラメータと別に調整すると、モデルの自由度を適度にあげられる場合が多い。

生物医学分野では、ハイパーパラメータと同様の決定方法によりカットオフの値を求めることが一般的である。C型慢性肝炎の肝線維化ステージ推定に関する従来研究では、カットオフの値の決め方を明記していないものが多いが [15], [17], [36]、この決定方法を取った可能性が高い。そこで本研究でも同じものを採用する。具体的には以下のとおりである。

カットオフを導入する場合、LOGR と KLOGR では2クラス問題用に式 (11) の置き換えを行ったうえで、分類規則を式 (2) から式 (12) の形にする。SVM の識別関数は符号によりクラスを表現するもので、すでに2クラス問題に特化しているため、式 (6) を式 (12) に直接用いる。MLR の識別関数は0か1の値を取る出力変数を予測するもので、

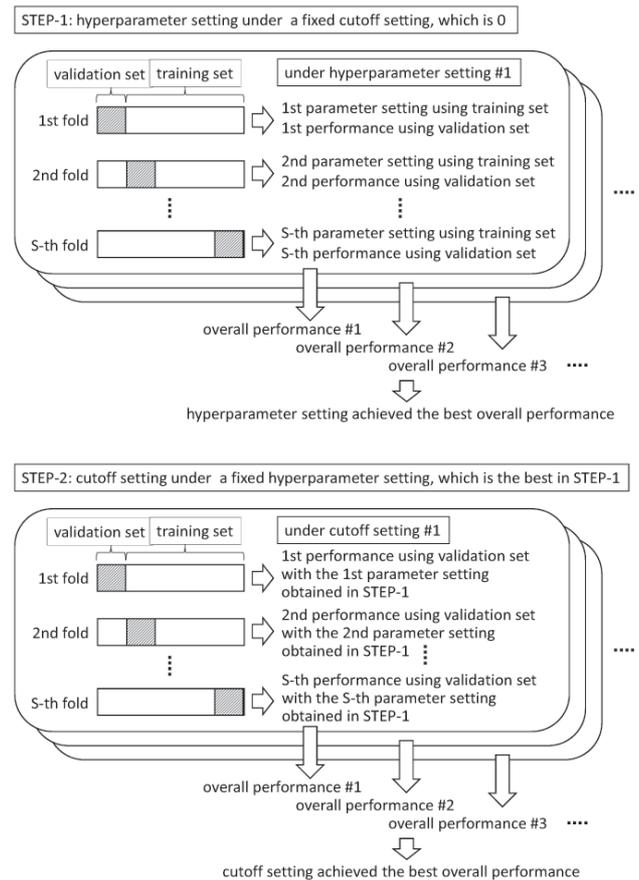


図 1 ハイパーパラメータとカットオフの設定方法

Fig. 1 Process of hyperparameter and cutoff setting.

これも2クラス問題特化であるため、SVMと同様である。

$$g_{-1}(\mathbf{x}) = p(C_{-1}|\mathbf{x}), \quad g_{+1}(\mathbf{x}) = p(C_{+1}|\mathbf{x}) \quad (11)$$

$$C(\mathbf{x}) = \begin{cases} C_{+1}, & \text{iff } g_{+1}(\mathbf{x}) - g_{-1}(\mathbf{x}) > cutoff \\ C_{-1}, & \text{iff } g_{+1}(\mathbf{x}) - g_{-1}(\mathbf{x}) \leq cutoff \end{cases} \quad (12)$$

カットオフ *cutoff* の設定には、 $cutoff_{lower} \leq cutoff \leq cutoff_{upper}$ の範囲をステップ $cutoff_{step}$ で探索するグリッドサーチを適用し、最も高い推定性能を達成した *cutoff* の値を見つける。識別関数 $g(\mathbf{x})$ がとりうる値の範囲に基づき、MLR では $0 \leq cutoff \leq 1$ 、LOGR, KLOGR, SVM では $-1 \leq cutoff \leq 1$ の探索範囲とする。探索ステップは共通して 0.01 を用いる。

ハイパーパラメータとカットオフを設定する一連の手続きを図 1 に示す。まず、STEP-1 ではカットオフを 0 に固定しておき、ハイパーパラメータ設定 #1 のもとに交差検定を行う。1~S の各フォールドにおいて（今回は Leave-One-Out 交差検定を使用するため、S は事例数に等しい）、訓練データで設定したパラメータの値と検証データで見積もった推定性能を求める。全フォールドの推定性能を総合することで、ハイパーパラメータ設定 #1 が持つ推定性能が得られる。グリッドサーチを適用してハイパー

表 2 本研究のデータセットに含まれる C 型慢性肝炎患者に関するデータを要約した結果. 年齢と検査値は中央値 (第 1 四分位-第 3 四分位), 性別は男女の人数の形式で表した. 肝線維化ステージの右側には事例数を記載した. なお, 従来, 慢性肝炎データを要約する際は F0 と F1 のデータを併合していたので, これに従った

Table 2 Summary of the data on chronic hepatitis C patients included in our dataset. It provides the median, first quartile, and third quartile of age, those of clinical test results, and the number of male and female patients. The number of instances is placed next to the corresponding liver fibrosis stage. The F0 and F1 stage data are merged in accordance with conventional ways.

Patient Information [Unit]	Fibrosis Stages			
	F0-F1:77	F2:28	F3:30	F4:17
AGE [year]	46.0 (35.8-56.3)	55.5 (52.5-59.0)	54.0 (49.0-58.0)	58.0 (50.3-61.3)
GENDER [num]	m:49, f:28	m:20, f:8	m:20, f:10	m:12, f:5

Clinical Tests [Unit]	Fibrosis Stages			
	F0-F1:77	F2:28	F3:30	F4:17
ALB [g/dL]	4.4 (4.2-4.7)	4.2 (4.1-4.6)	4.3 (4.1-4.5)	4.1 (3.7-4.2)
ALT [IU/l]	69.0 (31.5-139.3)	118.5 (80.0-210.5)	132.5 (98.0-202.0)	105.0 (70.0-167.3)
AST [IU/l]	46.0 (27.8-86.0)	76.0 (58.0-115.0)	97.0 (66.0-132.0)	89.0 (56.0-120.5)
CHE [IU/L]	297.0 (252.0-344.0)	261.5 (221.0-320.0)	249.0 (175.0-291.0)	232.0 (200.3-294.8)
F-A2G [%]	8.4 (7.7-9.4)	9.0 (8.4-9.7)	8.9 (8.3-9.8)	8.6 (7.3-9.6)
GGT [IU/l]	32.0 (18.0-65.3)	62.0 (27.5-88.0)	50.5 (33.0-86.0)	46.0 (35.3-77.3)
HGB [g/dl]	14.5 (13.6-15.4)	15.0 (13.5-15.7)	14.6 (13.8-15.6)	14.2 (13.5-14.9)
PLT [E03]	184.0 (158.0-222.5)	152.0 (129.0-173.5)	138.5 (116.0-183.0)	109.0 (87.0-141.0)
T-BIL [MG/DL]	0.7 (0.6-0.9)	0.7 (0.5-1.0)	0.8 (0.6-0.9)	0.9 (0.7-1.2)
T-CHO [MG/DL]	159.0 (142.8-178.5)	167.0 (153.0-200.0)	174.5 (152.0-186.0)	158.0 (145.3-169.5)
TP [G/DL]	7.4 (7.1-7.8)	7.4 (7.2-7.8)	7.6 (7.3-8.0)	7.3 (6.8-7.8)
TTT [U]	3.9 (1.9-6.7)	4.1 (3.3-7.3)	4.8 (3.1-7.9)	5.6 (1.6-7.6)
WBC [E03]	5.4 (4.6-6.3)	5.0 (4.3-6.0)	5.0 (4.0-6.7)	4.6 (3.8-5.5)
ZTT [U]	13.9 (10.6-17.7)	15.0 (12.9-20.8)	14.9 (12.9-18.6)	17.1 (11.7-21.5)

AGE: Age. GENDER: Gender, also called SEX. ALB: Albumin. ALT: Alanine aminotransferase, also called GPT. AST: Aspartate aminotransferase, also called GOT. CHE: Cholinesterase. F-A2G: Fecal α 2 globulin, also called F-A2.GL. GGT: γ -glutamyl transpeptidase, also called γ -GTP, GGTP, or G-GTP. HGB: Hemoglobin. PLT: Platelet count. T-BIL: Total bilirubin. T-CHO: Total cholesterol. TP: Total protein, also called T-PRO. TTT: Thymol turbidity test. WBC: White blood cell. ZTT: Zinc sulfate turbidity test.

パラメータ設定を変えながら本手続きを行うと, すべてのハイパーパラメータ設定の推定性能が出揃う. そして, この中で最も高い推定性能を達成したハイパーパラメータ設定を選択する.

STEP-2 ではハイパーパラメータを STEP-1 で求めた最適値に固定しておき, カットオフ設定#1 のもとに Leave-One-Out 交差検定を行う. 1~S の各フォールドにおいて, STEP-1 で設定済みのパラメータの値を用い, 検証データで見積もった推定性能を求める. そして, STEP-1 と同様の手続きを経て, 最も高い推定性能を達成したカットオフ設定を選択する. 最終的に, ハイパーパラメータ, パラメータ, カットオフのすべてに関して最適な設定とその推定性能が得られる.

3.3 公平な評価に向けた実験設計

副問題を解決して公平な評価を実現するために, 実験の詳細な設計を議論する. 副問題 1 の解決のために, 国際会議で公開された肝炎検査履歴データセット [5] を分析対象とする. 本データセットは, 千葉大学医学部附属病院で 1982~2001 年の 20 年間に収集されたものであり, 771 名の患者, 最大で 983 検査項目からなる. さらに詳しい情報に関しては, 国際会議 [5] や科研費プロジェクト [29] で提供された本データセットの説明書を参照されたい.

選定の一般性と再現性を確保すべく, 従来の主要な研究 [15], [17], [20], [36], [37] のそれぞれから検査項目の集合を得る. そして, 検査項目の集合の和集合を求め, さらにその和集合から一般的な通常の血液・尿検査に絞込む. 排他的な関係にある検査項目に関しては, 測定回数が最も多い検査項目 1 つを代表とする. 肝線維化ステージを知り

たい当日に実測値がない検査項目では、その1週間前（当日を含む）までに得られた最新の实測値で代用する。

結果として、ALB, ALT, AST, CHE, F-A2G, GGT, HGB, PLT, T-BIL, T-CHO, TP, TTT, WBC, ZTT という14検査項目を選定した。これらにAGE, GENDERの2つの患者基本情報を加えた全16項目を採用する。16項目の正式名称とともに、データを要約した結果を表2に示す。なお、慢性肝炎のうちC型に罹患した患者であり、16項目の全情報を持つという条件でデータを絞り込み、事例数は152個となった。

副問題1の解決には、分類器が学習対象とするパラメータの初期値を公平に設定することも重要である。本研究で扱う分類器はすべて凸な目的関数を持つため、非凸である場合の問題（パラメータの初期値依存性）は生じない。ゆえに、全分類器に共通してパラメータの初期値を0に設定することにした。

副問題2の解決のために、分類器間で異なる特徴量を用いていた従来研究[15], [17], [20], [36]とは異なり、我々は同じ項目から作成した特徴量を全分類器に共通で用いる。具体的には、14種類の検査項目と2種類の患者基本情報の値を平均0, 分散1に正規化したうえで並べ、1つの特徴量ベクトルにしたものをを入力する。

副問題3の根本的な解決には、従来使われてきた個々の評価基準や一部の評価基準の組合せではなく、評価基準のすべてを1つにまとめた評価基準が必要である。ゆえに我々は、全評価基準の算術平均 (Arithmetic Mean) と調和平均 (Harmonic Mean) を新規に導入し、これらを総合的な推定性能の評価基準とする [35]。算術平均は式 (13)、調和平均は式 (14) のように定義する。ただし、 C_i は i 番目の評価基準の値、 N_C は評価基準の数を意味する。

$AM(\{C_i \mid 1 \leq i \leq N_C\})$ は数値の大きさの差に対する感度が高く、チャンスレベルで50%となる。 $HM(\{C_i \mid 1 \leq i \leq N_C\})$ は数値の比率の差に対する感度が高く、0を1つでも含むと0%となる。たとえば、極端に少ない重症の事例を無視して全事例を軽症に分類すると、本研究のデータセットの場合、Accは肝線維化の深刻さの問題で51%、肝硬変の有無の問題で89%となるが、Sensはどちらも0%となる。 $HM(\{C_i \mid 1 \leq i \leq N_C\})$ は、このような場合に低い評価を与える。

$$AM(\{C_i \mid 1 \leq i \leq N_C\}) = \frac{1}{N_C} \sum_{i=1}^{N_C} C_i \quad (13)$$

$$HM(\{C_i \mid 1 \leq i \leq N_C\}) = \left(\frac{1}{N_C} \sum_{i=1}^{N_C} C_i^{-1} \right)^{-1} \quad (14)$$

2章の議論をふまえて、主要な5種類の評価基準からなる $AM = AM(\text{Acc}, \text{Sens}, \text{Spec}, \text{PPV}, \text{NPV})$, $HM = HM(\text{Acc}, \text{Sens}, \text{Spec}, \text{PPV}, \text{NPV})$ を総合的な推定性能の評価基準とする。この評価基準は、Acc, Sens, Spec, PPV,

NPVのすべてを重視する用途において、統一的な視点で性能を評価することを想定している。また、Accを除く4種類の評価基準からなる $AM' = AM(\text{Sens}, \text{Spec}, \text{PPV}, \text{NPV})$, $HM' = HM(\text{Sens}, \text{Spec}, \text{PPV}, \text{NPV})$ も評価基準として導入する。この評価基準は、Sens, Spec, PPV, NPVを重視し、Accの向上は問わない用途において、統一的な視点で性能を評価することを想定している。HMとHM'は評価基準間の値の差を抑制しつつ、すべての値を向上させる効果を持つので、ハイパーパラメータとカットオフの決定基準にも用いる。

推定性能の評価では、 $AM(\{C_i \mid 1 \leq i \leq N_C\}) = 50\%$, $HM(\{C_i \mid 1 \leq i \leq N_C\}) = 0\%$ を推定性能の下限値とする。まずは分類器がこの下限値を超えること、いい換えると、少なくとも推定に貢献することを確認する。そして、全分類器を推定性能の降順に並べ、高い推定性能を達成した分類器を調べる。汎化的な推定性能を厳密に見積もるため、 $AM(\{C_i \mid 1 \leq i \leq N_C\})$ と $HM(\{C_i \mid 1 \leq i \leq N_C\})$ の値の導出には、データセットの事例1つ1つに対してテストを行う Leave-One-Out 交差検定 [6] を採用する。

ただし、本研究のデータの事例数は多くはないため、訓練セット（パラメータの設定用）、検証セット（ハイパーパラメータとカットオフの設定用）、テストセット（完全に未知な事例に対する性能の見積り用）に3分割することは難しい。そこで今回は、Leave-One-Out 交差検定において取り出した1つの事例を検証とテストに併用する形をとった。

4. 評価実験

4.1 目的と条件

本実験の目的は、適用を提案する KLOGR と他の分類器の推定性能を比較して、KLOGR の有効性を検証することである。比較対象には、従来の主要な研究で適用された MLR [15], [26], [27], [36], LOGR [17], [26], [27], 1-NN [20], SVM [20], および、正則化ロジスティック回帰 (RLOGR: Regularized Logistic Regression) を用いる。RLOGR は LOGR を正則化したもの、KLOGR は LOGR を正則化・カーネル化したものである。RLOGR を比較対象に加えることで、KLOGR における正則化の影響とカーネル化の影響の区別を図る。カーネル法である KLOGR と SVM にはガウスクERNELを共通して用いる (表1参照)。

ハイパーパラメータの値の設定、パラメータの初期値の設定、学習におけるデータの与え方、収束基準、カットオフの値の設定は、3.2節で決定した手続きに従った。各設定の具体的な値を表3に示す。

対象問題は、2章で述べた2種類の2クラス問題とした。前処理、特徴量、推定性能の評価基準、推定システムは、3章で設計したとおりである。データセット [5] から3.3節で選定した16項目の全値を抽出し、152事例の個々について16項目からなる特徴量ベクトルを構成した。これを入

表 3 実験の条件設定. ハイパーパラメータの値, カットオフの値, パラメータの初期値, 収束基準の詳しい数値を示した. カットオフに関しては, 設定した値 (探索の下限値, 上限値, ステップ) を記載した. SVM のバイアス項は, カットオフのステップよりも高い精度で導出されるので, ここでは小数点以下 4 桁で記載した

Table 3 Experimental conditions. The details below are given: the values of hyperparameters and cutoff, the initial values of parameters, the criterion of convergence, and others related to optimization. A cutoff value is followed by its lower boundary, upper boundary, and step of search. The bias term of SVM has higher precision than the cutoff, and is rounded to four digits after the decimal point.

Classification of F0–F1/F2–F4					
Classifiers	Values of Hyperparameters and Cutoff	Initialization of Paramters	Maximum Number of Epocs	Learning Rate	Convergence Criterion
MLR	Cutoff = 0.52 (0, 1, 0.01)	0	None	None	None
LOGR	Cutoff = -0.08 (-1, 1, 0.01)	0	100000	0.001	0.0001
RLOGR	$\lambda = 1.85$ (0, 5, 0.01) Cutoff = 0.00 (-1, 1, 0.01)	0	100000	0.001	0.0001
1-NN	None	None	None	None	None
SVM (Bias) (Cutoff)	$\sigma = 3.56$ (0, 5, 0.01) $c = 1.49$ (0, 5, 0.01) Bias Term = -0.2629 Cutoff = -0.16 (-1, 1, 0.01)	0	100000 $\times N$	0.001	0.0001
KLOGR	$\sigma = 3.85$ (0, 5, 0.01) $\lambda = 3.44$ (0, 5, 0.01) Cutoff = 0.12 (-1, 1, 0.01)	0	100000	0.001	0.0001

Classification of F0–F3/F4					
Classifiers	Values of Hyperparameters and Cutoff	Initialization of Paramters	Maximum Number of Epocs	Learning Rate	Convergence Criterion
MLR	Cutoff = 0.21 (0, 1, 0.01)	0	None	None	None
LOGR	Cutoff = -0.84 (-1, 1, 0.01)	0	100000	0.001	0.0001
RLOGR	$\lambda = 3.00$ (0, 5, 0.01) Cutoff = -0.68 (-1, 1, 0.01)	0	100000	0.001	0.0001
1-NN	None	None	None	None	None
SVM (Bias) (Cutoff)	$\sigma = 4.83$ (0, 5, 0.01) $c = 1.07$ (0, 5, 0.01) Bias Term = 0.1261 Cutoff = 0.29 (-1, 1, 0.01)	0	100000 $\times N$	0.001	0.0001
KLOGR	$\sigma = 3.65$ (0, 5, 0.01) $\lambda = 0.26$ (0, 5, 0.01) Cutoff = -0.60 (-1, 1, 0.01)	0	100000	0.001	0.0001

力として分類器ごとに推定を行い, AM, AM', HM, HM' の値を得た. そして, $AM(\{C_i \mid 1 \leq i \leq N_C\}) = 50\%$, $HM(\{C_i \mid 1 \leq i \leq N_C\}) = 0\%$ という推定性能の下限値を超えるか, 推定性能の順位付けにおいて上位に位置するか, という観点から分類器を評価した. なお, 実験結果の一般性を保証するために, あらかじめ, 陰性と陽性のデータ分布を調べてステージ推定の難しさも見積もった.

4.2 結果と考察

陰性と陽性のデータ分布を調べた結果, 肝線維化の深刻

さの問題では, 陰性 (F0–F1) は 77 個, 陽性 (F2–F4) は 75 個, 陰性に対する陽性の比率は 0.9740 であった. 肝硬変の有無の問題では, 陰性 (F0–F3) は 135 個, 陽性 (F4) は 17 個, 陰性に対する陽性の比率は 0.1259 であった. 陰性よりも陽性が極端に少ない傾向は, 従来研究 [15], [17], [20], [36] における傾向と同様である. ただし, 本研究のデータセットの方が陽性の比率が低く, 従来研究よりもステージ推定がやや難しいと考えられる.

表 4 にステージ推定の結果を示す. 表 4 の上側は, 肝線維化の深刻さの 2 クラス問題において分類器の推定性能

表 4 実験の結果. 上側は肝線維化の深刻さの 2 クラス問題, 下側は肝硬変の有無の 2 クラス問題で得られたもの. 推定性能の降順に分類器を並べた. SVM (Bias) は SVM 本来のバイアス項を用いた結果, SVM (Cutoff) はカットオフを導入した結果である

Table 4 Experimental Results. The upper section corresponds to the binary-class classification problem on the seriousness of fibrosis, and the lower on the existence of cirrhosis. The classifiers are in the descending order of prediction performance. SVM (Bias) and SVM (Cutoff) mean SVM using a bias term and a cutoff, respectively.

Classification of F0-F1/F2-F4					
Classifiers	AM	AM'	Classifiers	HM	HM'
KLOGR	85.54 (1st)	85.54 (1st)	KLOGR	85.53 (1st)	85.53 (1st)
SVM (Cutoff)	77.78 (2nd)	77.81 (2nd)	SVM (Cutoff)	77.60 (2nd)	77.59 (2nd)
SVM (Bias)	77.24 (3rd)	77.30 (3rd)	SVM (Bias)	76.89 (3rd)	76.87 (3rd)
RLOGR	75.21 (4th)	75.26 (4th)	MLR	74.96 (4th)	74.95 (4th)
MLR	75.14 (5th)	75.17 (5th)	RLOGR	74.93 (5th)	74.91 (5th)
LOGR	73.97 (6th)	74.04 (6th)	LOGR	73.56 (6th)	73.53 (6th)
1-NN	64.48 (7th)	64.48 (7th)	1-NN	64.47 (7th)	64.47 (7th)
Lower Boundary	50.00	50.00	Lower Boundary	0.00	0.00

Classification of F0-F3/F4					
Classifiers	AM	AM'	Classifiers	HM	HM'
KLOGR	73.53 (1st)	70.20 (1st)	KLOGR	66.35 (1st)	62.65 (1st)
SVM (Bias)	72.15 (2nd)	69.63 (2nd)	SVM (Bias)	63.23 (2nd)	59.78 (2nd)
SVM (Cutoff)	71.59 (3rd)	68.27 (3rd)	SVM (Cutoff)	63.09 (3rd)	59.28 (3rd)
RLOGR	67.82 (4th)	66.36 (4th)	RLOGR	55.00 (4th)	51.72 (4th)
LOGR	66.27 (5th)	65.07 (5th)	LOGR	52.45 (5th)	49.23 (5th)
MLR	61.67 (6th)	58.83 (6th)	MLR	45.90 (6th)	42.00 (6th)
1-NN	58.77 (7th)	52.58 (7th)	1-NN	27.94 (7th)	23.95 (7th)
Lower Boundary	50.00	50.00	Lower Boundary	0.00	0.00

を求め, 降順に並べた結果である. 全分類器が下限値を大幅に超え, 少なくとも推定に対する効果が見られた.

KLOGR は AM, AM', HM, HM' のすべてで 1 番目に高い推定性能を示した. SVM は, バイアス項を用いた条件とカットオフを用いた条件の間でかなり近い結果となり, これらは AM, AM', HM, HM' のすべてで 2 番目に高い推定性能を示した. KLOGR は SVM (推定性能が高いカットオフ使用の結果を選択) に比べて, AM が 7.76%, AM' が 7.73%, HM が 7.93%, HM' が 7.94% 高く, 推定性能に差が見られた.

また, KLOGR は従来最も広く使われてきた MLR や LOGR に比べて AM が 10.40% 以上, AM' が 10.37% 以上, HM が 10.57% 以上, HM' が 10.58% 以上高く, 推定性能に顕著な差が見られた. ゆえに, 肝線維化の深刻さの問題では, KLOGR が最も高い推定性能を持つこと, SVM が次に続くことが分かった.

表 4 の下側は, 肝硬変の有無の 2 クラス問題において分類器の推定性能を求め, 降順に並べた結果である. 全分類器が下限値を大幅に超え, 少なくとも推定に対する効果が見られた.

KLOGR は AM, AM', HM, HM' のすべてで 1 番目に

高い推定性能を示した. SVM は, バイアス項を用いた条件とカットオフを用いた条件の間でかなり近い結果となり, これらは AM, AM', HM, HM' のすべてで 2 番目に高い推定性能を示した. KLOGR は SVM (推定性能が高いバイアス項使用の結果を選択) に比べて, AM が 1.38%, AM' が 0.57%, HM が 3.12%, HM' が 2.87% 高く, 推定性能に差が見られた.

また, KLOGR は従来最も広く使われてきた MLR や LOGR に比べて AM が 7.26% 以上, AM' が 5.13% 以上, HM が 13.90% 以上, HM' が 13.42% 以上高く, 推定性能に顕著な差が見られた. ゆえに, 肝硬変の有無の問題では, KLOGR が最も高い推定性能を持つこと, SVM が次に続くことが分かった. 以上より, 2 種類の 2 クラス問題に共通して KLOGR が最も有効であり, SVM がその次に有効であるといえる.

4.3 総合的な分析と評価

ここでは実験データをまとめて次の 3 つの分析を行い, KLOGR を SVM と対比しながら考察を行う. 分析 1: 総合的な推定性能の順位を導出し, KLOGR の有効性の安定性を調べる. 分析 2: AM, AM', HM, HM' を構成する推

表 5 実験における分析 1 の結果. 2 つの 2 クラス問題のデータを総合し, 分類器を平均推定性能の降順に示した

Table 5 Results of Analysis #1. The prediction performances obtained in two types of binary-class classification problems are averaged and provided in descending order.

Average for Two Classification Problems					
Classifiers	AM	AM'	Classifiers	HM	HM'
KLOGR	79.53 (1st)	77.87 (1st)	KLOGR	77.99 (1st)	76.17 (1st)
SVM (Bias)	74.69 (2nd)	73.47 (2nd)	SVM (Bias)	72.63 (2nd)	71.07 (2nd)
SVM (Cutoff)	74.69 (2nd)	73.04 (3rd)	SVM (Cutoff)	70.35 (3rd)	68.44 (3rd)
RLOGR	71.52 (4th)	70.81 (4th)	RLOGR	68.83 (4th)	67.58 (4th)
LOGR	70.12 (5th)	69.55 (5th)	LOGR	67.18 (5th)	66.00 (5th)
MLR	68.40 (6th)	67.00 (6th)	MLR	65.50 (6th)	63.67 (6th)
1-NN	61.63 (7th)	58.53 (7th)	1-NN	55.50 (7th)	52.23 (7th)
Lower Boundary	50.00	50.00	Lower Boundary	0.00	0.00

定性能の評価基準の値を導出し, KLOGR の各評価基準に対する改善効果を調べる. 分析 3: 分析 1, 2 の結果に基づき, KLOGR と SVM が他の分類器よりも推定性能が高い理由, および, KLOGR が SVM よりも推定性能が高い理由を検討する.

分析 1

KLOGR の有効性の安定性を検証するため, 以下の手続きにて総合的な推定性能の順位を求めた. 肝線維化の深刻さ, 肝硬変の有無に関する 2 つの 2 クラス問題で得られたデータを用い, 分類器ごとに全実験条件の AM の平均を計算した. AM', HM, HM' についても同様に平均を計算した. 表 5 に分類器を総合的な推定性能の降順に並べた結果を示す. KLOGR は AM, AM', HM, HM' のすべてで 1 番目に高い推定性能を達成した. 特に MLR や LOGR と比較すると, KLOGR は MLR よりも AM が 11.13%, AM' が 10.87%, HM が 12.49%, HM' が 12.50% 高く, LOGR よりも AM が 9.41%, AM' が 8.32%, HM が 10.81%, HM' が 10.17% 高い. 従来最もよく使われてきた MLR や LOGR よりも, KLOGR の方がかなり有効であることが分かる. KLOGR の順位は, 肝線維化の深刻さの問題, 肝硬変の有無の問題の両方において, AM, AM', HM, HM' のすべてでつねに 1 位であった.

以上より, KLOGR は今回用いた分類器の中で最も推定に有効であり, その有効性は安定しているといえる. 加えて, KLOGR ではクラスの事後確率を求め, これを推定の確信度の目安とすること, 多クラスへの拡張性を生かして肝線維化ステージをより細かく分類することも可能である.

SVM は, バイアス項を用いた条件とカットオフを用いた条件の間でかなり近い結果となり, これらは AM, AM', HM, HM' のすべてで 2 番目に高い推定性能を示した. SVM の順位は, 肝線維化の深刻さの問題, 肝硬変の有無の問題の両方において, AM, AM', HM, HM' のすべてでつねに 2 位であった. ゆえに総合的にも, SVM が KLOGR

表 6 実験における分析 2 の結果. 上側は肝線維化の深刻さの 2 クラス問題, 下側は肝硬変の有無の 2 クラス問題で得られたもの. Acc, Sens, Spec, PPV, NPV の各々に関して, KLOGR, SVM が達成した値, 全分類器の最小値と最大値を求めた. SVM (Bias) は SVM 本来のバイアス項を用いた結果, SVM (Cutoff) はカットオフを導入した結果である

Table 6 Results of Analysis #2. The upper section corresponds to the binary-class classification problem on the seriousness of fibrosis, and the lower on the existence of cirrhosis. For each of Acc, Sens, Spec, PPV, and NPV, the following are provided: The values achieved by KLOGR, those by SVM, and the maximum and minimum values by all the classifiers. SVM (Bias) and SVM (Cutoff) mean SVM using a bias term and a cutoff, respectively.

Classification of F0-F1/F2-F4					
Classifiers	Acc	Sens	Spec	PPV	NPV
KLOGR	85.53	86.67	84.42	84.42	86.67
SVM (Bias)	76.97	84.00	70.13	73.26	81.82
SVM (Cutoff)	77.63	82.67	72.73	74.70	81.16
Maximum Value	85.53	86.67	84.42	84.42	86.67
Minimum Value	64.47	65.33	63.64	63.64	65.33

Classification of F0-F3/F4					
Classifiers	Acc	Sens	Spec	PPV	NPV
KLOGR	86.84	52.94	91.11	42.86	93.89
SVM (Bias)	82.24	64.71	84.44	34.38	95.00
SVM (Cutoff)	84.87	52.94	88.89	37.50	93.75
Maximum Value	86.84	70.59	92.59	42.86	95.24
Minimum Value	71.05	11.76	71.11	16.67	89.29

に次ぐ推定性能を持つことが示された.

分析 2

ここでは, AM, AM', HM, HM' を構成する推定性能の評価基準に対して, KLOGR, および, SVM が持つ改善効果を検証する. 表 6 に, 肝線維化の深刻さ, 肝硬変

表 7 実験における分析 2 の結果. 2 つの 2 クラス問題のデータを総合し, Acc, Sens, Spec, PPV, NPV の各々に関して, KLOGR, SVM が達成した値, 全分類器の最小値と最大値を求めた

Table 7 Results of Analysis #2. For each of Acc, Sens, Spec, PPV, and NPV, the values obtained in two types of binary-class classification problems are averaged, and the following are provided: The average values achieved by KLOGR, those by SVM, and the maximum and minimum average values by all the classifiers.

Average for Two Classification Problems					
Classifiers	Acc	Sens	Spec	PPV	NPV
KLOGR	86.18	69.80	87.76	63.64	90.28
SVM (Bias)	79.61	74.35	77.29	53.82	88.41
SVM (Cutoff)	81.25	67.81	80.81	56.10	87.46
Maximum Value	86.18	75.96	87.76	63.64	90.28
Minimum Value	72.37	38.55	68.67	40.15	77.31

の有無の個々の 2 クラス問題で得られた各評価基準の値 (KLOGR の値, SVM の値, 全分類器の最大値, 全分類器の最小値) を示す. また, 総合的な傾向を見るため, 2 つの 2 クラス問題で得られたデータを用い, 分類器ごとに全実験条件の Acc, Sens, Spec, PPV, NPV のそれぞれの平均を計算した. 表 7 に, 各評価基準の平均値 (KLOGR の値, SVM の値, 全分類器の最大値, 全分類器の最小値) を示す.

KLOGR は, 表 6 の上側ではすべての推定性能の評価基準で最大値を持ち, 下側では Sens を除き, 最大値かそれに近い値を持っている. 表 7 の総合結果を見ると, KLOGR では, Sens の値は最大値よりも 6.16%低い, 他のすべての推定性能の評価基準で最大値を達成している. よって, C 型慢性肝炎の肝線維化ステージ推定というタスクでは, KLOGR には全評価基準の値をバランス良く高める効果があるといえる. ただし, Sens が最大値よりも低い点, すなわち, 陽性的見落としがある点については, 今後, 改善策を考える必要がある. 一方, SVM では, バイアス項を用いた条件, カットオフを用いた条件の両者でほぼ最大値に次ぐ値を達成しているが, 評価基準間のバランスをやや欠くように見える.

分析 3

分析 1, 2 の結果から, C 型慢性肝炎の本データセットでは KLOGR と SVM は他の分類器よりも推定性能が高いこと, および, KLOGR は SVM よりも推定性能が高いことが明らかになった. 前者の理由は, KLOGR と SVM がカーネル化により非線形な識別境界を柔軟に表現できるから, と考えられる.

ただし, カーネル化には正則化をとまなうため, 正則化が功を奏したかもしれない. そこで, LOGR, LOGR に正則化を施した RLOGR, LOGR に正則化とカーネル化を施し

た KLOGR を比較してみる. 表 4, 表 5 より, RLOGR は LOGR と推定性能が近く, KLOGR は LOGR や RLOGR よりも大幅に推定性能が高い. ゆえに, KLOGR の高い推定性能は主にカーネル化に起因するといえる. 今回, SVM に正則化のみを施した分類器は用いなかったが, SVM に関しても同様であると考えられる.

次に, KLOGR と SVM の間で推定性能の差異が生じた理由を検討する. これらの間で, カーネル化と正則化, ハイパーパラメータとカットオフの設定方法, 目的関数の第 2 項は共通しており, 目的関数の第 1 項が異なる. SVM の第 1 項はヒンジ誤差関数, KLOGR の第 1 項は交差エントロピー誤差関数であり, 文献 [6] の Fig. 7.5 のグラフ (横軸: 誤分類尺度, 縦軸: 誤差関数) にプロットされた形状を持つ. 識別境界付近における 2 つの誤差関数の形状の違いが性能の差異を生じた可能性が考えられる. 特に, 誤差関数の正分類の領域に着目すると, 正分類の場合, SVM では誤差は 0 となり学習が進まないが, KLOGR では誤差は小さいが 0 ではなく学習が進む, という違いがある.

ただし, 正分類でも学習が進むからといって, クラスの事後確率を推定する KLOGR の目的関数が最良とは限らない. クラスの事後確率の推定では, クラス境界付近に位置して Acc~NPV の値を左右する重要な事例に対する感度が, むしろ低下するからである. 本研究では KLOGR の優位性が示されたが, さらに高いステージ推定の性能を得るためには, KLOGR の利点を生かしつつ, クラス境界付近の事例への感度が高い目的関数を導入する改善が望まれる.

5. おわりに

過去に, 重回帰 (MLR), ロジスティック回帰 (LOGR), サポートベクターマシン (SVM) を適用し, 血液・尿の検査値から C 型慢性肝炎の肝線維化ステージを推定する研究がなされてきた. しかし, MLR や LOGR には高い推定性能を得られないという問題があり, SVM にはクラスの事後確率が得られない, 多クラスへの拡張が難しいという問題があった. そこで本研究では, LOGR と SVM の両方の利点を持つカーネルロジスティック回帰 (KLOGR) に着目し, その適用を提案した. 実験の結果, KLOGR は SVM 等の他の分類器よりも推定性能が高かった. 特に, 従来最もよく使われてきた MLR や LOGR に比べて, KLOGR は 8.32~12.50%も高い推定性能を達成した. KLOGR には, クラスの事後確率の導出, 多クラスへの拡張の容易さという利点もあり, 総合的に見て KLOGR は有効であるといえる.

参考文献

[1] Altman, D.G. and Bland, J.M.: Diagnostic Tests 1: Sensitivity and Specificity, *British Medical Journal*, Vol.308, p.1552 (1994).

- [2] Altman, D.G. and Bland, J.M.: Diagnostic Tests 2: Predictive Values, *British Medical Journal*, Vol.309, p.102 (1994).
- [3] Altman, D.G. and Bland, J.M.: Diagnostic Tests 3: Receiver Operating Characteristic Plots, *British Medical Journal*, Vol.309, p.188 (1994).
- [4] Balasubramanian, V., Ho, S.-S. and Vovk, V.: *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations and Applications*, Elsevier, Amsterdam, Holland (2014).
- [5] Berka, P. and Cremilleux, B.: ECML/PKDD Discovery Challenge 2005 (online), available from (<http://lisp.vse.cz/challenge/ecmlpkdd2005/>) (accessed 2014-12-31).
- [6] Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer, New York (2006).
- [7] Castera, L., Vergniol, J., Foucher, J., et al.: Prospective Comparison of Transient Elastography, Fibrotest, APRI, and Liver Biopsy for the Assessment of Fibrosis in Chronic Hepatitis C, *Gastroenterology*, Vol.128, No.2, pp.343–350 (2005).
- [8] Cawley, G.C. and Talbot, N.L.C.: Efficient Model Selection for Kernel Logistic Regression, *Proc. IEEE International Conference on Pattern Recognition ICPR-2004*, Vol.2, pp.439–442 (2004).
- [9] Cornfield, J.: Joint Dependence of Risk of Coronary Heart Disease on Serum Cholesterol and Systolic Blood Pressure: A Discriminant Function Analysis, *Federation Proc., Federation of American Societies for Experimental Biology*, No.11, pp.58–61 (1962).
- [10] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol.20, pp.273–297 (1995).
- [11] Crammer, K. and Singer, Y.: On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines, *Journal of Machine Learning Research*, Vol.2, pp.265–292 (2001).
- [12] Cristianini, N. and Shawe-Taylor, J.: *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, Cambridge, England (2000).
- [13] Deng, L. and Li, X.: Machine Learning Paradigms for Speech Recognition: An Overview, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.21, No.5, pp.1060–1089 (2013).
- [14] Fakhraei, S., Soltanian-Zadeh, H., Fotouhi, F., et al.: Confidence in Medical Decision Making: Application in Temporal Lobe Epilepsy Data Mining, *Proc. 2011 Workshop on Data Mining for Medicine and Healthcare*, pp.60–63 (2011).
- [15] Forns, X., Ampurdanes, S., Llovet, J.M., et al.: Identification of Chronic Hepatitis C Patients without Hepatic Fibrosis by a Simple Predictive Model, *Hepatology*, Vol.36, No.4, pp.986–992 (2002).
- [16] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York (2009).
- [17] Imbert-Bismut, F., Ratziu, V., Pieroni, L., et al.: Biochemical Markers of Liver Fibrosis in Patients with Hepatitis C Virus Infection: A Prospective Study, *The Lancet*, Vol.357, pp.1069–1075 (2001).
- [18] 泉 並木: ガイドライン/ガイドンス慢性肝炎こう診る・こう考える, 日本医事新報社, 東京 (2011).
- [19] Jaakkola, T., Meila, M. and Jebara, T.: *Maximum Entropy Discrimination*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Massachusetts, USA (1999).
- [20] Jiang, Z., Yamauchi, K., Yoshioka, K., et al.: Support Vector Machine-Based Feature Selection for Classification of Liver Fibrosis Grade in Chronic Hepatitis C, *Journal of Medical Systems*, Vol.30, pp.389–394 (2006).
- [21] Koizumi, Y., Hirooka, M., Kisaka, Y., et al.: Liver Fibrosis in Patients with Chronic Hepatitis C: Noninvasive Diagnosis by Means of Real-time Tissue Elastography – Establishment of the Method for Measurement, *Radiology*, Vol.258, No.2, pp.610–617 (2011).
- [22] 厚生労働省: C型肝炎対策等に関する専門家会議報告書 C型肝炎対策等の一層の推進について, 厚生労働省 (オンライン), 入手先 (<http://www.mhlw.go.jp/houdou/2005/08/dl/h0802-2b.pdf>) (参照 2014-12-31).
- [23] Lee, M.H., Cheong, J.Y., Um, S.H., et al.: Comparison of Surrogate Serum Markers and Transient Elastography (Fibroscan) for Assessing Cirrhosis in Patients with Chronic Viral Hepatitis, *Digestive Diseases and Sciences*, Vol.55, pp.3552–3560 (2010).
- [24] Lewin, M., Poujol-Robert, A., Boelle, P.-Y., et al.: Diffusion-Weighted Magnetic Resonance Imaging for the Assessment of Fibrosis in Chronic Hepatitis C, *Hepatology*, Vol.46, No.3, pp.658–665 (2007).
- [25] Matsuda, K., Ohsaki, M., Katagiri, S., et al.: Application of Kernel Logistic Regression to the Prediction of Liver Fibrosis Stages, *Proc. International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems SCIS-ISIS02012*, pp.780–784 (2012).
- [26] Mauss, S., Berg, T., Rockstroh, J., et al.: *Hepatology, A Clinical Textbook, Edition 2012*, Flying Publisher (2012).
- [27] McDonald, J.W.D., Burroughs, A.K., Feagan, B.G., et al.: *Evidence-based Gastroenterology and Hepatology, Third Edition*, Wiley-Blackwell, Hoboken, New Jersey (2010).
- [28] Memisevic, R.: *Dual Optimization Conditional Probability Models*, University of Toronto, Toronto, Canada (2006).
- [29] 元田 浩: 共通データ解析報告書科学研究費特定領域研究「情報洪水時代におけるアクティブマイニングの実現」(オンライン), 入手先 (<http://www.ar.sanken.osaka-u.ac.jp/activemining/final.pdf>) (参照 2014-12-31).
- [30] Nouretdinov, I., Costafreda-Gonzalez, S., Gammernan, A., et al.: Machine learning classification with confidence: Application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression, *NeuroImage*, Vol.56, No.2, pp.809–813 (2011).
- [31] 小俣政男: Bed Side ノートシリーズ肝炎, 現代医療社, 東京 (2001).
- [32] Roth, V.: Probabilistic Discriminative Kernel Classifiers for Multi-class Problems, *Lecture Notes in Computer Science*, Vol.2191, pp.246–253 (2001).
- [33] Stefanescu, H., Grigorescu, M., Lupsor, M., et al.: A New and Simple Algorithm for the Noninvasive Assessment of Esophageal Varices in Cirrhotic Patients Using Serum Fibrosis Markers and Transient Elastography, *Journal of Gastrointestinal and Liver Diseases*, Vol.20, No.1, pp.57–64 (2010).
- [34] Tanaka, K., Kurita, T. and Kawabe, T.: Selection of Import Vectors via Binary Particle Swarm Optimization and Cross-Validation for Kernel Logistic Regression, *Proc. IEEE International Joint Conference on Neural Networks IJCNN-2007*, pp.1037–1042 (2007).
- [35] van Rijsbergen, C.J.: *Information Retrieval*, Butter-

worths (1979).

- [36] Wai, C.-T., Greenson, J.K., Fontana, R.J., et al.: A Simple Noninvasive Index Can Predict Both Significant Fibrosis and Cirrhosis in Patients with Chronic Hepatitis C, *Hepatology*, Vol.38, No.2, pp.518-526 (2003).
- [37] Yamada, Y., Suzuki, E., Yokoi, H., et al.: Experimental Evaluation of Time-Series Decision Tree, *Lecture Notes in Artificial Intelligence*, Vol.3430, pp.190-209 (2005).
- [38] Zhu, J. and Hastie, T.: Kernel Logistic Regression and the Import Vector Machine, *Journal of Computational and Graphical Statistics*, Vol.14, No.1, pp.185-205 (2005).



大崎 美穂 (正会員)

1994年九州芸術工科大学(現、九州大学)芸術工学部音響設計学科卒業。1996年同大学院芸術工学研究科博士前期課程修了。1999年博士後期課程修了。静岡大学情報学部助手、同志社大学工学部専任講師を経て、現在、同志社大学理工学部准教授。博士(工学)。IEEE, 人工知能学会, 知能情報ファジィ学会, 音響学会各会員。



松田 健司

2011年同志社大学工学部情報システムデザイン学科卒業。2013年同大学院理工学研究科博士前期課程修了。現在、ヤフー株式会社勤務。IEEE会員。



ワン ベン

2011年西安電子科技大学(中国)情報科学技術学院卒業。現在、西安電子科技大学・同志社大学ダブルディグリープログラム博士前期課程在学。



片桐 滋 (正会員)

1977年東北大学工学部電気工学科卒業。1979年同大学院工学研究科修士課程修了。1982年同大学院工学研究科博士課程修了。日本電信電話公社(現、日本電信電話株式会社)、(株)国際電気通信基礎技術研究所を経て、現在、同志社大学理工学部教授。工学博士。IEEE, 日本音響学会, 電子情報通信学会各会員。



横井 英人

1996年香川医科大学医学部卒業。同年千葉大学医学部附属病院第一内科にて臨床研修。千葉大学医学部附属病院医療情報部医員, 厚生労働省勤務, 香川大学医学部附属病院医療情報部講師を経て、現在、香川大学医学部附属病院医療情報部教授, 臨床研究支援センター長。医学博士。日本医療情報学会, 日本内科学会, 日本消化器内視鏡学会, 日本超音波医学会, 日本遠隔医療学会各会員。



高林 克日己

1975年千葉大学医学部卒業。同年千葉大学医学部附属病院第二内科にて臨床研修。国保旭中央病院医師, 国保松戸市立病院医長, 千葉大学医学部附属病院講師, 松戸市立福祉医療センター東松戸病院内科部長, 千葉大学医学部附属病院教授を経て、現在、医療法人社団鼎会三和病院顧問。医学博士。日本医療情報学会, 日本内科学会, 日本呼吸器学会, 日本リウマチ学会, 日本老年医学会, 日本医学教育学会各会員。