

放射線治療のための深層学習モデルに基づく 呼吸性移動予測の時系列データ拡張による性能向上

Performance improvement of deep-learning-based prediction of respiratory motion by using data augmentation for radiation therapy

○石井 万結*、市地 慶*、篠原 唯*、淡路 樹*、本間 経康*

○Mayu Ishii*, Kei Ichiji*, Yui Shinohara*, Itsuki Awaji*, Noriyasu Homma*

*東北大学大学院医学系研究科

*Tohoku University Graduate School of Medicine

キーワード: 放射線治療 (Radiation therapy)、肺がん (Lung cancer)、深層学習 (Deep learning)、Transformer、データ拡張 (Data augmentation)

連絡先: 〒980-8575 仙台市青葉区星陵 2-1 東北大学大学院医学系研究科 医用画像工学分野

石井 万結, Tel.: 022-717-8190, E-mail: mayu.ishii.q3@dc.tohoku.ac.jp

1. はじめに

平均寿命延伸に伴い、わが国では生涯のがんの罹患率は50%を超え、男性ではおよそ4人に1人、女性では6人に1人ががんにより死亡するとされる[1]。がんの治療は、罹患部位・臓器、進行状況、全身状態などをもとに治療方針を決定し、外科的療法、化学療法、放射線治療といった治療法を単独もしくは組み合わせることで進められる。主要な治療法のうち放射線治療は低侵襲かつ副作用が局所的であることから、とくに肺や肝臓などの外科的療法では機能が失われる可能性がある臓器にも適用しやすい、高齢者などの体力が低下している患者でも治療を受けやすいなどの利点が知られている[2]。

放射線治療では、確実な治療効果を得るために治療対象である標的腫瘍へと放射線を十分に照射すると同時に、副作用低減のために治療対象周辺の正常組織の被ばくを

極力抑制することの両立が求められる。しかし、体幹部の腫瘍は体内臓器の影響を受けるため、その位置が時々刻々と変化することが知られている。とくに呼吸の影響を受けやすい肺がんでは、標的腫瘍が呼吸によって時々刻々と位置変化する[3]。このような呼吸性移動を伴う腫瘍へ十分な線量投与し、周辺の肺野の被ばく線量を低減するためには、呼吸性移動への対策が必要であり[4]、最大吸息時や最大呼息時など特定の呼吸相にあるときのみ照射する呼吸同期照射や、呼吸性移動に応じて照射範囲を追従させる動体追尾照射が開発・利用されている。

呼吸性移動に応じて照射を制御する対策法は、腫瘍位置の呼吸による移動をリアルタイムに計測し、時空間的に標的腫瘍のみへ投与線量を集中することを実現する[4]。しかし、位置計測や照射制御の双方に必要な処理・駆動により、計測されたタイミング

と照射制御が完了するまでには合わせて数十～数百ミリ秒程度の時間遅れが発生すると報告されている[5]。この時間遅れは標的腫瘍の範囲と照射の範囲とにずれを生じさせる。そこで、このずれを補償するために、呼吸性移動による将来の腫瘍位置変動の時系列予測を行い、照射を先回りして制御することが求められている[6]。

呼吸性移動予測手法として、これまでに統計的時系列モデル、機械学習モデル、深層学習モデルを用いたさまざまな予測手法が提案されており、近年では深層学習モデルによる呼吸性移動予測が優れていると期待されている[7]。深層学習を用いた呼吸性移動予測の研究として、再帰型ニューラルネットワーク (Recurrent Neural Network: RNN) の一種であるゲート付き再帰型ユニット (Gated recurrent unit: GRU)[8]を用いたものが報告されている[9]。RNN は自身の持つ再帰型構造により、過去の情報をそれ以降の時点の出力算出にも活かされるという特徴があり、時系列予測を含む系列処理に優れている。また近年では系列の要素ごとの関係を学習することができる注意機構 (Attention) を用いた深層学習モデルである Transformer[10]に基づくモデル群が自然言語処理の分野において優れた性能を達成している[11]。さらに時系列予測に特化したモデルである Informer[12]による呼吸性移動予測が従来手法と比べ優れていたとの報告がある[13]。

しかし深層学習による呼吸性移動予測には呼吸の性質変化があるデータでの予測が適切に行うことができないという課題がある。図1はGRU、およびInformerによる呼吸性移動予測の一例である。5秒から15秒の付近をみると、黄色線及び黄緑線で示

したGRU、Informerによる予測結果は、黒線で示される真値とかけ離れた予測結果になっていることがわかる。この図1の例において、予測モデルの学習および評価に使用したデータ[14]を図2に示す。図2中の破線は訓練データ、評価データそれぞれの平均値を示している。それぞれの平均値に差があり、訓練データと評価データ間で呼吸の性質変化があったとみなせる。深層学習モデルの予測が適切に行うことができない原因は、このような性質の差にモデルが対応できていないことだと考えられる。そのため、深層学習モデルの予測性能を十分に引き出すためには、将来起こり得る呼吸の性質変化をも予測モデルに学習させる必要がある。訓練データを元にバリエーションに富んだデータを作成し、それを学習に用いるデータ拡張は有効であると考えられる。

以上より、本研究では、深層学習モデルによる呼吸性移動予測において、呼吸の性質変化を含むデータでの予測性能を向上させることを目的として、呼吸性移動時系列向けのデータ拡張の効果を検証する。訓練用の呼吸性移動時系列データにみられる主要な性質変化を模擬することで、より多様な呼吸パターンを含む訓練データへと拡張する。拡張した訓練データをもとにTransformerベースモデルであるInformerモデルを構築し、呼吸性移動予測におけるデータ拡張を実験的に評価する。

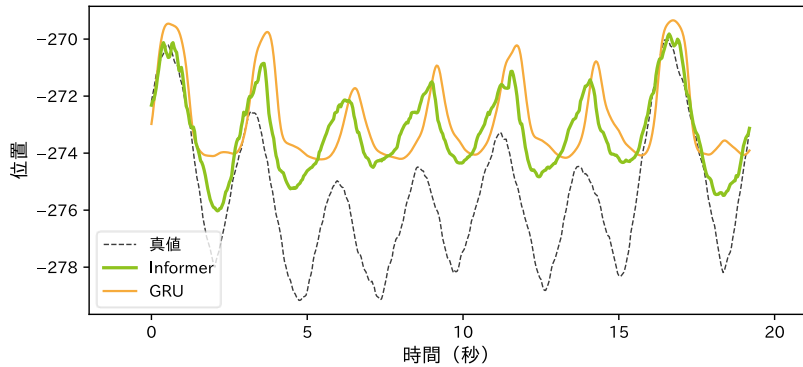


図 1: GRU、Informer による呼吸性移動予測の一例

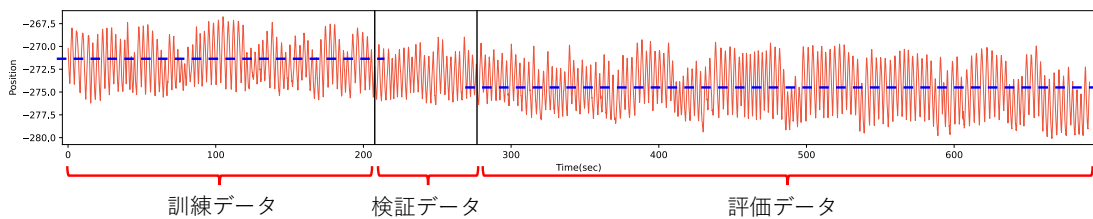


図 2: 図 1 の予測モデルの学習および評価に使用したデータ

2. 時系列データ拡張を用いた深層学習モデルによる呼吸性移動予測

本章では、呼吸性移動予測における時系列データ拡張の効果を検証するために、時系列向けの Transformer ベースモデルである Informer モデルにデータ拡張を適用する方法について説明する。

2.1. Transformer に基づく呼吸性移動予測モデル

2.1.1. Transformer

Transformer は主に注意機構 (Attention) で構成されるエンコーダ-デコーダ型モデルである。注意機構 (Attention) は、系列の要素ごとの関係を学習する機構である。同じく系列処理向けとして知られる RNN と比べてより系列内の長期の関係性を扱うことに優れている。しかし、注意機構自体には順序関係の概念が存在しないため、要素同士の位置関係を捉えるために位置情報の埋め込みを行う。

2.1.2. Informer: 本研究の使用モデル

Informer は Transformer を時系列予測向けに改良したモデルである。Transformer での位置情報の埋め込みの代わりに、時系列に含まれている秒、分、時間、日、週、月、年や曜日といった時間情報の埋め込みを行う。しかし呼吸性移動のデータには日、曜日などの概念はない。呼吸性移動予測向けには、通常的时间情報に加えて吸息や呼息といった呼吸の位相に関する情報の埋め込みを行う。呼吸の位相としては、呼吸性移動予測時系列の瞬時位相を用いる。

瞬時位相はヒルベルト変換を用いて求めることができる。時刻 t における呼吸性移動時系列 $x(t)$ の瞬時位相 $\theta(t)$ は以下の式 (1) で表される。

$$\theta(t) = \tan^{-1}\left(\frac{H[x(t)]}{x(t)}\right) \quad (1)$$

ただし、 $x(t)$ は時刻 t における腫瘍位置、 $H[x(t)]$ は $x(t)$ のヒルベルト変換である。

次に線形層 $f(\cdot)$ を通じて腫瘍位置と瞬時

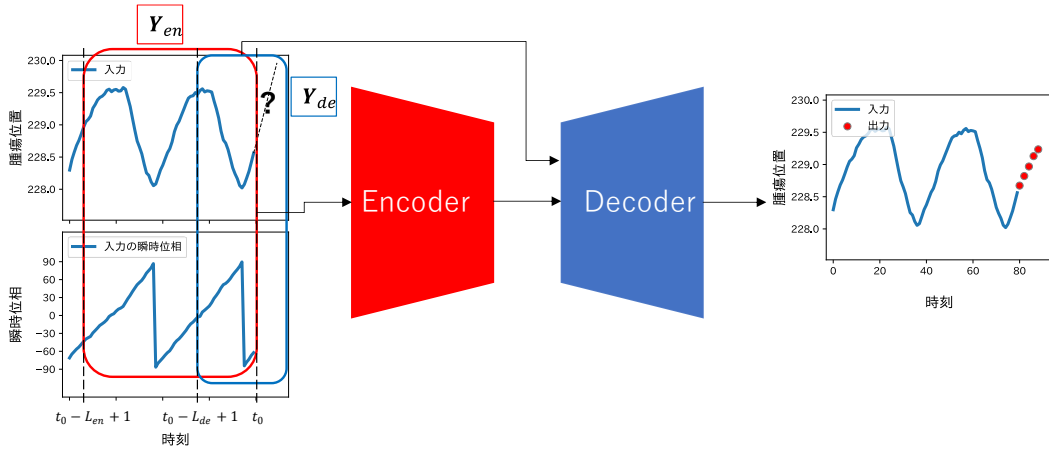


図 3: 本研究使用モデルの入出力の関係

位相を高次元特徴ベクトルへと変換し、それらを合成した特徴ベクトル $\mathbf{y}(t)$ を求める。

$$\mathbf{y}(t) = \mathbf{f}(x(t), \theta(t)) \quad (2)$$

本研究で用いたモデルの入出力の関係を図3に示す。図3中のエンコーダ入力 \mathbf{Y}_{en} 、デコーダ入力 \mathbf{Y}_{de} は、特徴ベクトル $\mathbf{y}(t)$ を用いて、以下の式(3)、式(4)で表すことができる。

$$\mathbf{Y}_{en} = \begin{bmatrix} \mathbf{y}(t_0 - L_{en} + 1) \\ \mathbf{y}(t_0 - L_{en} + 2) \\ \vdots \\ \mathbf{y}(t_0) \end{bmatrix} \quad (3)$$

$$\mathbf{Y}_{de} = \begin{bmatrix} \mathbf{y}(t_0 - L_{de} + 1) \\ \mathbf{y}(t_0 - L_{de} + 2) \\ \vdots \\ \mathbf{y}(t_0) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad (4)$$

2.2. 呼吸性移動時系列のデータ拡張

図1で示したように深層学習による呼吸性移動予測が適切に行うことができなかったデータの性質変化には、i) 呼吸周期の変化、そしてii) 重力およびほかの臓器の影響による腫瘍位置の呼吸よりも緩やかな変動(以下、長期変動)の変化が含まれている。本節では、呼吸周期の変化および長期変動の変化に対応するための2つの拡張手法

について説明する。

2.2.1. 時間方向の伸縮

呼吸周期の変化に対応するためのデータ拡張として、時間方向の伸縮を行う。これにより、様々な呼吸周期を持つ呼吸性移動データを模擬することができる。時刻 t における腫瘍位置を $x(t)$ 、変換後の時刻および腫瘍位置をそれぞれ t' 、 $x'(t')$ としたとき、以下の式(5)で示される変換を行うことにより、時間方向の伸縮を実現する。

$$\begin{cases} t' = mt \\ x'(t') = x(t) \end{cases} \quad (5)$$

ただし、 m は時間方向伸縮の係数であり、一様分布 $U(0.9, 1.1)$ に従う。

図4に $m > 1$ であるときの、時間方向の伸縮の様子を示す。ある時刻 t_1 における腫瘍位置 $x(t_1)$ が、 $t'_1 = mt_1$ における腫瘍位置 $x'(t'_1)$ に変換されていることがわかる。また時間方向への引き伸ばしにより、呼吸周期の異なったデータが作成されたことがみてとれる。

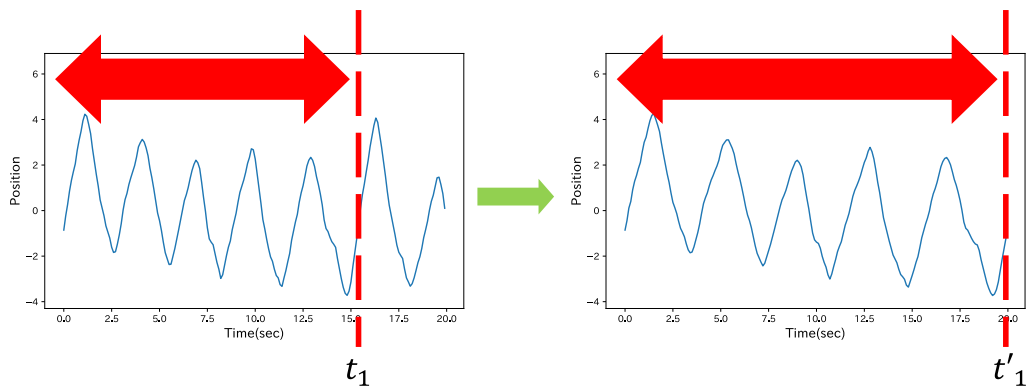


図 4: 時間方向の伸縮

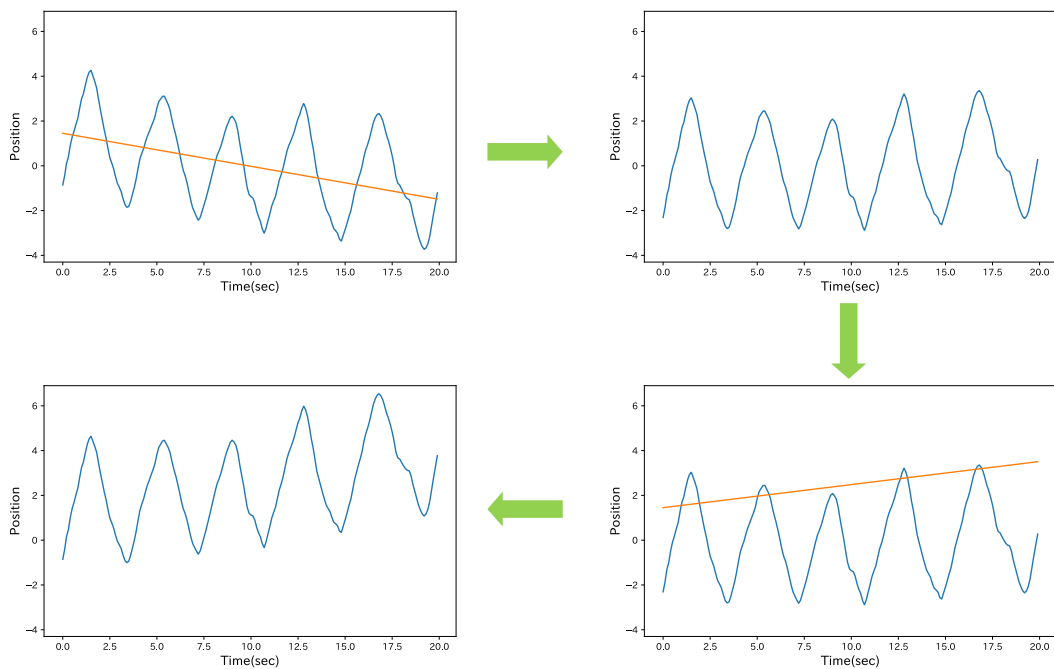


図 5: 長期変動の変更

2.2.2. 長期変動の変更

重力およびほかの臓器の移動による腫瘍位置の長期変動の変化に対応するために、将来起こり得る長期変動を模擬するように長期変動の傾きを変更する。本研究では図 5 左上の橙線のように長期変動を一次近似直線で表すことにする。

長期変動の変更では、まず時間方向の伸縮を行った後の呼吸性移動時系列 $x'(t')$ から最小二乗法により求めた元の長期変動 (図 5 左上橙線) を取り除く。時刻 t' における腫瘍位置 $x'(t')$ から元の長期変動を取

り除いた腫瘍位置 $x_{NoTrend}(t')$ は式(6)で表される。

$$x_{NoTrend}(t') = x'(t') - (at' + b) \quad (6)$$

ただし、 a 、 b は時間方向の伸縮を行った後腫瘍位置 $x'(t')$ の一次近似直線の式の係数である。図 5 右上に元の長期変動を取り除いた $x_{NoTrend}(t')$ の例を示す。

次に取り除いた長期変動の傾きである a を基準として図 5 右下橙線のような新たな長期変動を作成し、それを加えることで図 5 左下のように元データと異なる長期変動をもつデータを作成する。長期変動変更後

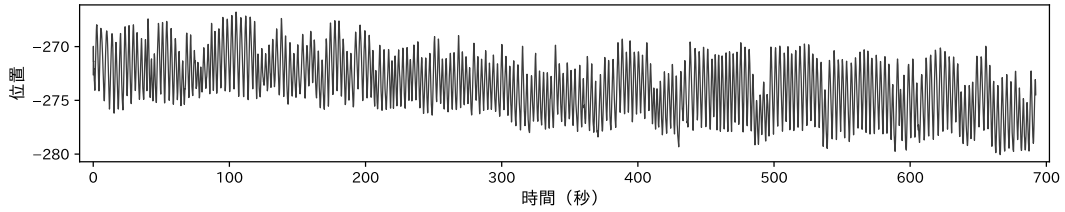


図 6: 本研究で使したデータの一例

の腫瘍位置 $x''(t')$ は式(7)で表される。

$$x''(t') = x_{NoTrend}(t') + (lat' + b) \quad (7)$$

ただし、 l は長期変動変更の係数であり、一様分布 $U(-1.1, +1.1)$ に従う。

3. 検証実験

呼吸性移動予測向けの時系列データ拡張を用いた Informer モデルの予測性能を評価するため、呼吸性移動の実データを用いて予測実験を行った。

3.1 使用データ

呼吸性移動予測において広く用いられているオープンデータ[14]のうち、15人分を使用した。これらのデータには性質変化の影響により深層学習での予測が適切に行えなかったデータを含んでいる。予測が適切に行えなかったデータの一例を図 6 に示す。サンプリング周波数 26 Hz のデータを 10 Hz にダウンサンプリングし使用した。データ長は 692.3 秒(6923 点)であり、訓練 30%、検証 10%、評価 60%に分割した。

3.2 比較対象モデル

本研究では、データ拡張の効果を調べるため、①Informer モデルにデータ拡張を適用したモデル (以下、データ拡張適用モデル) と、②データ拡張を適用しなかったモデル (以下、データ拡張未適用モデル) の比較を行った。データ拡張の有無以外はすべて同条件で実験を行った。①データ拡張適用モデルでは、時間方向の伸縮範囲は 0.9~1.1 倍 ($m \sim U(0.9, 1.1)$)、長期変動の傾きの

範囲は、基準の傾きの -1.1 ~ 1.1 倍 ($l \sim U(-1.1, 1.1)$) とした。なお学習と評価はデータごとに 10 試行した。本研究で使したハイパーパラメータを表 1 に示す。ハイパーパラメータは②データ拡張未適用モデルでのグリッドサーチをもとに決定し、①データ拡張適用モデルと共通とした。

③予測による遅れ補償を行わずに照射したときに相当する零時ホールド法による予測 (以下、予測による遅れ補償なし) も行い、比較対象とした。

$$\hat{x}(t+h) = x(t) \quad (8)$$

なお $\hat{x}(t+h)$ は時刻 $t+h$ における予測値、 $x(t)$ は時刻 t における実測値である。予測先 h はすべてのモデルで 0.5 秒 (5 点) とした。

3.3 実験環境

Informer モデルは Python (ver.3.9.7) 環境下で、深層学習フレームワーク PyTorch (ver.1.10.0) を用いて実装した。実験には、CPU: Intel Core i9-10900K、メモリ: 64 GB、GPU: NVIDIA RTX A6000 の計算機を使用した。

3.4 評価指標

予測結果の評価指標として、Root mean squared error (RMSE) を用いた。RMSE は式(9)により求められる。

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(t) - \hat{x}(t))^2} \quad (9)$$

RMSE は小さい値になるほど予測が正確であることを示している。

4. 検証実験の結果

データ 2 およびデータ 13 における予測結果の例をそれぞれ図 7、図 8 に示す。また、データ 2 およびデータ 13 における RMSE をそれぞれ表 2、表 3 に示す。図 7 の緑枠線内に着目すると、②データ拡張未適用モデルが腫瘍位置の低下に対応できていないのに対し、①データ拡張適用モデルは低下した部分に追従した予測を行うことができている。また、表 2 より①データ拡張適用モデルの RMSE は、②データ拡張未適用モデルおよび③予測による遅れ補償なしよりも小さくなっていることがわかる。以上から、データ 2 において①データ拡張適用モデルは②データ拡張未適用モデルよりも真値に近い予測ができたといえる。

一方、図 8 の緑枠線内に着目すると、①データ拡張適用モデルは呼吸の谷の部分で真値よりも低い値に予測してしまい、真値との差が大きくなっていることがわかる。表 3 から、RMSE による比較でも①データ

拡張適用モデルが②データ拡張未適用モデルよりも大きくなっていることが分かり、データ拡張を行ったことにより予測精度が低下したといえる。

図 9 に、10 試行分の RMSE の分布をデータごとに示す。また表 4 にデータごとの RMSE の 10 試行平均を示す。図 9 を見ると、多くのデータで①データ拡張適用モデルの RMSE が②データ拡張未適用モデルよりも小さくなっていることがわかる。また①データ拡張適用モデルでの予測が③予測による遅れ補償なしよりも予測精度が低下することがなくなったことがわかる。一方データ 5、データ 11、データ 13 ではデータ②拡張未適用モデルの RMSE が①データ拡張適用モデルよりも下回っているように見受けられる。表 4 より、RMSE の 10 試行平均は、15 データのうち 11 データで①データ拡張適用モデルが最小であることがわかる。しかし 15 データのうち 4 データでは②データ拡張未適用モデルの RMSE が①データ拡張適用モデルよりも小さくなっている。

表 1: 本実験で用いたハイパーパラメータ

| パラメータ | 値 | パラメータ | 値 |
|-------------------|-----|---------|--------|
| 位相計算用の入力長 | 100 | モデル内次元 | 512 |
| Encoder 入力長 | 40 | 全結合層次元 | 4096 |
| Decoder 入力長 | 10 | バッチ数 | 32 |
| Decoder 出力長 | 5 | 最大エポック数 | 500 |
| Attention head の数 | 8 | 学習率 | 0.0001 |
| Encoder のレイヤー数 | 2 | ドロップアウト | 0.05 |
| Decoder のレイヤー数 | 1 | 損失関数 | MSE |
| 入力次元&出力次元 | 1 | 最適化手法 | Adam |

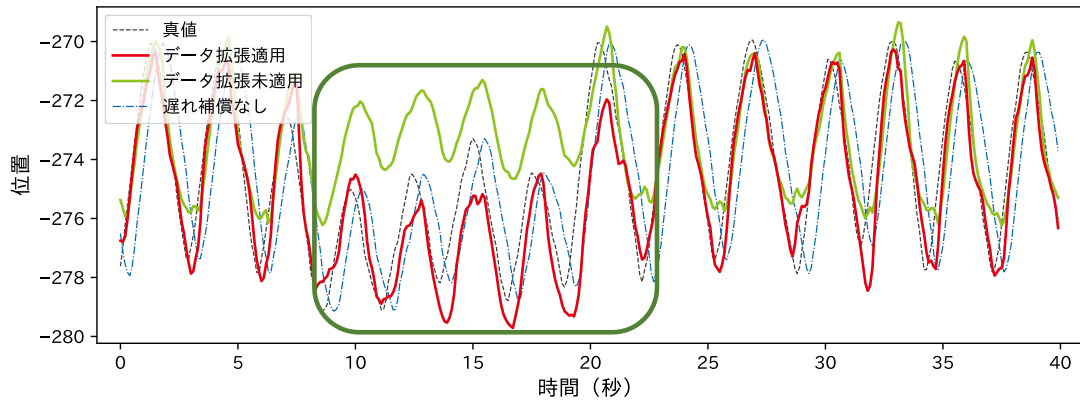


図 7: データ 2 における予測

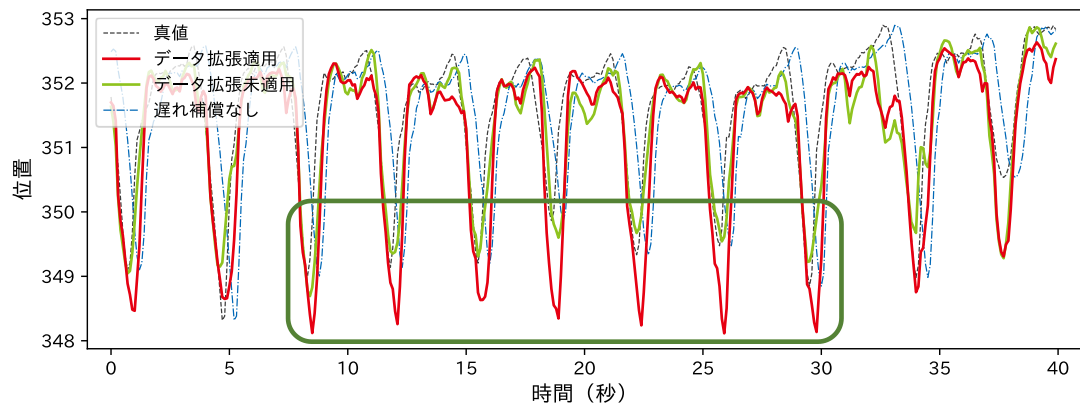


図 8: データ 13 における予測

表 2: データ 2 における予測の RMSE

| モデル | RMSE |
|--------------|-------|
| ①データ拡張適用モデル | 0.965 |
| ②データ拡張未適用モデル | 2.25 |
| ③予測による遅れ補償なし | 2.10 |

表 3: データ 13 における予測の RMSE

| モデル | RMSE |
|--------------|-------|
| ①データ拡張適用モデル | 0.646 |
| ②データ拡張未適用モデル | 0.559 |
| ③予測による遅れ補償なし | 1.22 |

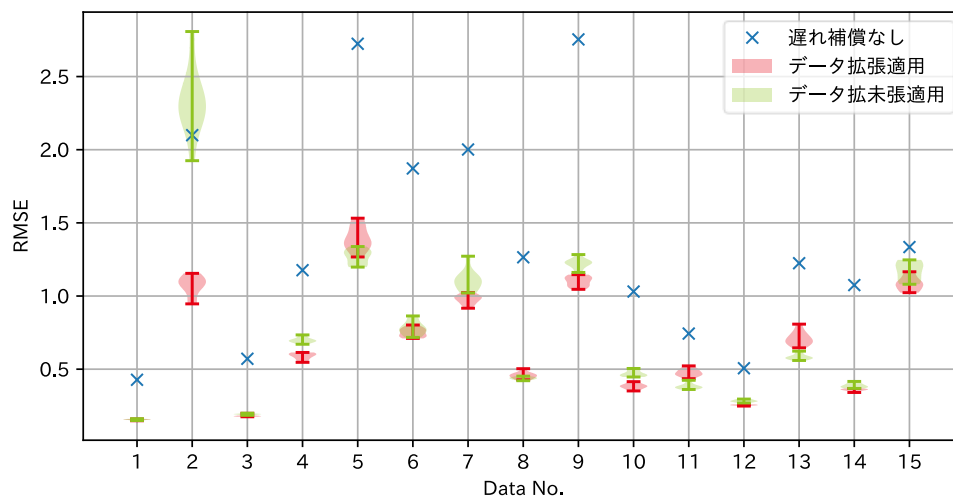


図 9: データごとの 10 試行分の RMSE の分布

表 4: データごとの RMSE10 試行平均

| データ番号 | ①データ拡張 適用モデル | ②データ拡張 未適用モデル | ③予測による 遅れ補償なし |
|--------|-----------------|------------------|------------------|
| データ 1 | 0.155 | 0.156 | 0.427 |
| データ 2 | 1.07 | 2.32 | 2.10 |
| データ 3 | 0.181 | 0.192 | 0.571 |
| データ 4 | 0.589 | 0.699 | 1.18 |
| データ 5 | 1.39 | 1.27 | 2.72 |
| データ 6 | 0.751 | 0.779 | 1.87 |
| データ 7 | 0.977 | 1.11 | 2.00 |
| データ 8 | 0.475 | 0.438 | 1.26 |
| データ 9 | 1.10 | 1.22 | 2.75 |
| データ 10 | 0.383 | 0.470 | 1.03 |
| データ 11 | 0.477 | 0.382 | 0.743 |
| データ 12 | 0.255 | 0.283 | 0.507 |
| データ 13 | 0.710 | 0.584 | 1.22 |
| データ 14 | 0.358 | 0.387 | 1.07 |
| データ 15 | 1.08 | 1.17 | 1.33 |

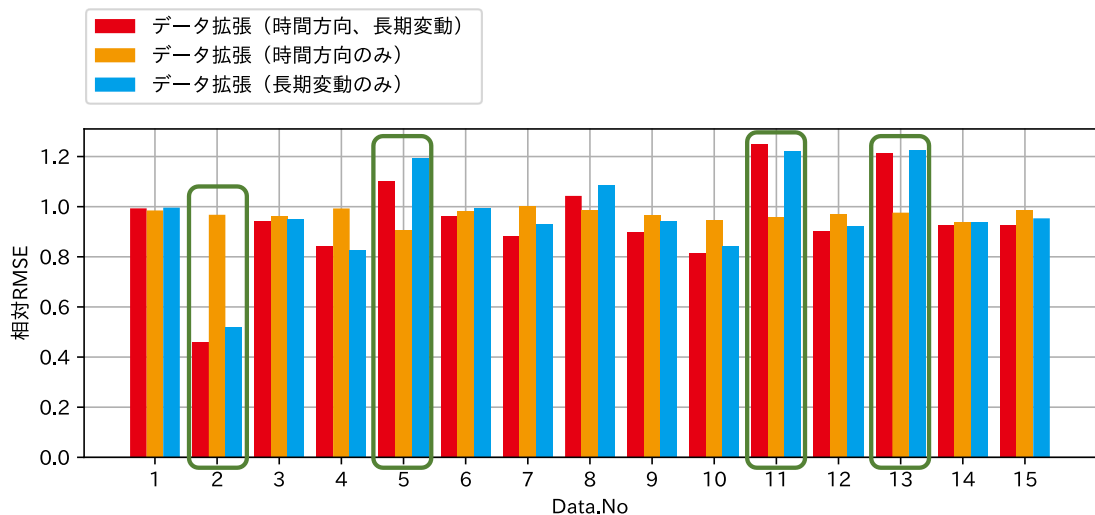


図 10: データ拡張未適用モデルを基準とした RMSE を 1 としたときの各モデルの相対化 RMSE

5. 考察

検証実験の結果より、15 データのうち 11 データで、データ拡張による性能向上が見られた。これは Informer モデルを用いた予測においてデータ拡張が有効である可能性を示唆している。一方、データ 2 やデータ 5、データ 13 ではデータ拡張による予測性能低下が見られた。このことはデータ拡張がモデルに悪影響を与えることがあることを示している。とくに、本研究では時間方向の伸縮と、腫瘍位置の長期変動の変更、2 種類の拡張方法を適用したが、どちらの拡張方法がどの程度性能向上・低下に影響を与えたのかについても特定が望まれる。そこで、時間方向の伸縮のみ適用したモデル、長期変動の変更のみ適用したモデルを訓練し、予測性能の追加評価を行った。ハイパーパラメータは 3.2 と同様に設定し、各データ 10 回ずつ学習と予測を行った。

図 10 に(a) 2 種類のデータ拡張適用モデル、(b)時間方向の伸縮のみ適用したモデル、

(c)長期変動の変更のみ適用したモデルの 10 試行の平均 RMSE について、データ拡張未適用モデルを基準として相対化した値を示す。この相対化 RMSE が 1 未満であればデータ拡張による性能改善があったこと意味し、1 を超えている場合には予測性能に悪影響があったことを意味する。図 10 の緑枠線内に注目すると、データ 2 では(a)と(c)の相対化 RMSE が 0.5 以下となっており、長期変動の変更により大きな改善が生じたことがわかる。一方、データ 5、データ 11、データ 13 を見ると同じく(a)と(c)の相対化 RMSE が 1 を超えており。これらデータでは長期変動の変更の適用による悪影響が生じていたことがわかる。また(a)と(b)の相対化 RMSE より、時間方向の伸縮はデータ拡張未適用モデルからの性能低下こそ生じないものの性能向上への寄与はそれほど大きくないことがわかる。また、全体としてデータにより相対化 RMSE の傾向が異なることから、拡張方法の種類ごとの適用可否はデータごとに向き不向きがあることがうか

がえる。したがって、今後どのようなデータにおいて予測性能低下がどのように生じるのか、その傾向や原因をさらに検討する必要がある。

このほか本研究の2種類の拡張方法で対応しきれていない呼吸性変動の性質が存在する可能性がある。このため、ほかに考慮すべき性質の調査とその拡張方法の実現手段の検討も必要である。

6. おわりに

本研究では、放射線治療の動体追尾照射を想定した、深層学習モデルによる呼吸性移動予測に時系列データ拡張を適用し、その効果を検証した。データ拡張は、時系列向け深層学習モデルであるInformerモデルに、呼吸周期の変化に対応するための時間方向の伸縮、重力およびほかの臓器の移動による腫瘍位置の長期的変動の変化に対応するための長期変動の変更を適用した。

データ拡張を適用したモデルによる予測は、データ拡張を適用しないモデルと比較して、多くのデータで予測精度の改善がみられた。このことから深層学習モデルによる呼吸性移動予測においてデータ拡張が有効である可能性が示唆された。その一方、一部データにおいてデータ拡張を適用したことによる性能低下がみられた。またそれぞれのデータ拡張方法はデータによって向き不向きが存在した。そのため、データにより適用するデータ拡張方法および条件を変更する必要があると考えられる。

今後の課題として、データ拡張により予測性能が低下した例に対する改善策を検討することが挙げられる。本研究で用いたデータ拡張方法は、性能向上、低下どちらにもつながることがあるため、改善策を講じる

ことにより、より多くのデータでの性能向上を実現することが望まれる。

また適切なデータ拡張方法と拡張条件の検討も課題である。より多くの呼吸性変化に対応できるデータ拡張方法を適用することにより、さらなる性能向上につなげることが期待される。

参考文献

- [1] がんの統計編集委員会, *がんの統計〈2022年版〉*. 公益財団法人がん研究振興財団, 2022.
- [2] 小塚隆弘、稲邑清也, 診療放射線技術下巻, 改訂第14版. 南江堂, 2019.
- [3] Y. Suh, S. Dieterich, B. Cho, and P. J. Keall, “An analysis of thoracic and abdominal tumour motion for stereo-tactic body radiotherapy patients,” *Phys Med Biol*, vol. 53, no. 13, pp. 3623–3640, Jul. 2008, doi: 10.1088/0031-9155/53/13/016.
- [4] P. J. Keall *et al.*, “The management of respiratory motion in radiation oncology report of AAPM Task Group 76a),” *Med Phys*, vol. 33, no. 10, pp. 3874–3900, Oct. 2006, doi: 10.1118/1.2349696.
- [5] P. R. Poulsen, B. Cho, A. Sawant, D. Ruan, and P. J. Keall, “Detailed analysis of latencies in image-based dynamic MLC tracking,” *Med Phys*, vol. 37, no. 9, pp. 4998–5005, Sep. 2010, doi: 10.1118/1.3480504.
- [6] P. J. Keall *et al.*, “AAPM Task Group 264: The safe clinical implementation of MLC tracking in radiotherapy,” *Med Phys*, vol. 48, no. 5, pp. e44–e64, May 2021, doi: 10.1002/MP.14625.
- [7] A. Mylonas, J. Booth, and D. T. Nguyen, “A review of artificial intelligence applications for motion tracking in radiotherapy,” *J Med Imaging Radiat Oncol*, vol. 65, no. 5, pp. 596–611, Aug. 2021, doi: 10.1111/1754-9485.13285.
- [8] K. Cho *et al.*, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,”

- EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734, Jun. 2014, doi: 10.3115/v1/d14-1179.
- [9] S. Yu, J. Wang, J. Liu, R. Sun, S. Kuang, and L. Sun, “Rapid Prediction of Respiratory Motion Based on Bidirectional Gated Recurrent Unit Network,” *IEEE Access*, vol. 8, pp. 49424–49435, 2020, doi: 10.1109/ACCESS.2020.2980002.
- [10] A. Vaswani *et al.*, “Attention Is All You Need,” *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 5999–6009, Jun. 2017, doi: 10.48550/arxiv.1706.03762.
- [11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2018, doi: 10.48550/arxiv.1810.04805.
- [12] H. Zhou *et al.*, “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting,” *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, vol. 12B, pp. 11106–11115, Dec. 2020, doi: 10.48550/arxiv.2012.07436.
- [13] 西山亮、市地慶、本間経康、杉田典大, “注意機構に基づく肺腫瘍位置変動予測,” 第55回日本生体医工学会東北支部大会, 2021.
- [14] “Signals @ ROB,” https://signals.rob.uni-luebeck.de/index.php?title=Signals_@_ROB.